

International panel for GM food?

A proposal to set up a body analogous to the Intergovernmental Panel on Climate Change has much to commend it. But it raises questions that should critically influence the allocation of responsibility.

It would have been unrealistic to expect any dramatic conclusions from last week's three-day meeting on the safety of genetically modified (GM) foodstuffs, hosted by the UK government and run jointly with the Organisation for Economic Co-operation and Development (OECD; see page 112). Conceived at the height of vocal demands last year for a global moratorium on the commercial planting of GM foods, its main function was to help lower the temperature of a debate that was rapidly leading to a trade war. In this it appears to have succeeded. But many of the gaps between those keen to promote such technologies and their critics remain as wide as ever.

Any attempt to establish a permanent forum in which the dialogue can continue, as proposed at the end of the meeting, could be ineffectual for those reasons. Broad differences in philosophy remain — for example, between those convinced that GM crops hold the key to meeting the food requirements of the Third World, and others who say they are unnecessary because the problem is primarily one of food distribution. Such differences will not be resolved by a discussion whose starting point is the science behind safety issues. They require a broader agenda and more mutual trust between participants — or at least more common ground over the facts and practicalities.

Nor should one be seduced by the simplistic argument that what worked for climate change will also work for GM crops. The Intergovernmental Panel on Climate Change was able to organize much of its work around a single question — whether there is a man-made component of current global warming — to which many scientists already felt they had, if not the answer, then a great deal to contribute. 'Is GM food safe?' is a very different type of question, requiring

elements of judgement that cannot be resolved through data and simulations alone.

More positively, an international panel could present political decision-makers with conclusions shorn of the more excessive claims of either the proponents or critics of GM foods. Thus, it could play a useful role in identifying the key issues at stake, for example on the implications for human health or environmental safety.

Achieving this does not require starting from the science, even though scientific judgements as to which claims are more plausible than others would be a major component in such a panel's deliberations, and could help establish important new scientific agendas to sustain public confidence. An equally important component would be the trust in the process held by stakeholders in the debate. Here, openness, transparency and a commitment to priority consideration of the interests not only of commerce but also of consumers worldwide are crucial. So is the willingness to acknowledge the legitimacy of a wide range of views, not just to listen to them politely. Any organization keen to take on responsibility for such a panel would do well to commit itself publicly to such requirements.

In offering itself as a candidate, the OECD has a lot to prove. The proceedings of an important conference held in 1997 on the controversial regulatory issue of substantial equivalence have only recently been de-restricted, not least because of the time it took to satisfy participating governments about the report's contents. That sort of track record, and a historical lack of time for anti-GM lobbyists, does not, on the face of it, bode well for the OECD's suitability to take responsibility for the stewardship of this highly charged debate. ■

Re-reforming Unesco

The organization's director-general should focus more closely on core goals.

History suggests that Koïchiro Matsuura, director-general of the United Nations Educational, Scientific and Cultural Organization (Unesco) since late last year, has a few years in which to improve it before he risks being taken in by its own propaganda and becomes as inward-looking as his most recent predecessors. A forthright speech last week (see page 113) shows that he means well. Only time will tell whether he can change the culture so radically that he avoids the self-aggrandizing traps that snared Amadou-Mahtar M'Bow and, less disastrously but still notably, Federico Mayor.

Following the fiefdom that Unesco became in the '70s and '80s under M'Bow, which led to the exit of the United States, Singapore and the United Kingdom, Mayor achieved significant reforms, restoring sufficient confidence to encourage the United Kingdom to rejoin. But latterly he lost that touch. Last year's World Conference on Science was widely seen as having failed to make the most of a significant opportunity to place science at the centre of the international political and economic stage.

The success of Matsuura — a Japanese lawyer, diplomat and deputy foreign minister — will depend on whether he is able to entice the United States back into the organization. This will at least need a

prompt but fair examination of the questionable 'eleventh-hour' appointments Mayor sprang on Unesco just before leaving. Commandingly, Matsuura is forthright in highlighting the difficulties he faces, and is rightly focusing first on management.

But a sharpening of Unesco's programmes is essential. Some have suggested that it should take a lead in the International Year for the Culture of Peace. A look at Unesco's declaration on its Culture of Peace programme is not encouraging — a plethora of worthy motions spanning the sale of small arms and the rights of women. Like most of the motions carried at the World Conference on Science, these have little to offer by way of leverage and, therefore, seem likely to do little for the future reputation of Unesco or for world peace.

At the heart of Unesco's mandate is a commitment to education, with science and culture inextricably linked with it. Science education is not the answer to many short-term problems in society's handling of science-related issues, but its improvement is an urgent need nevertheless, in both developed and developing countries. Some prompt, tangible new achievements to that end, within existing programmes and also embedded as goals in its next medium-term strategy, could do Unesco a lot of good. ■





G8 on GM
Krebs' panel urges industrial nations to get it together
p112



Waste of space?
Zero-gravity crystals have yet to prove their worth
p114



AIDS therapy
Panel ponders South Africa's policy on AZT and other drugs
p115



Military fallout
India's research budget gains from political tension
p116

Genome leaders told to keep their eyes on the main prize

JAMES D WILSON/LIAISON

Washington

The Human Genome Project is in danger of being sidetracked from its goal of sequencing our entire genetic blueprint by 2003. That is the fear of several prominent advisers to the project, who are worried that enthusiasm for turning the emerging 'rough draft' of the human genome into a finished product is waning.

Their concern emerged last week at a meeting of the US National Advisory Council for Human Genome Research. It was sparked by a surge of interest in initiatives designed to make sense of the draft human genome. These include efforts to sequence the mouse genome and to develop computational tools and databases.

The Human Genome Project has sequenced about 1.7 billion unique base pairs so far. By June, the international coalition of publicly funded and charity-supported labs should have sequenced another billion. This will give a draft covering 90 per cent of the genome, sequenced five times over.

Finishing the job means filling hard-to-sequence gaps and ensuring that 99 per cent of the genome has been sequenced ten times over — the 'gold standard' to ensure that the sequence is accurate. But once geneticists have their hands on the draft, this goal may lose its appeal. At least in the short term, many are more interested in partial information on single genes than a complete genome.

This is where the other initiatives come in. Because mouse genes can be experimentally 'knocked out', the mouse genome can help to define the function of genes common to both mouse and human. Creating a library of mammalian complementary DNA (cDNA), representing the coding sequences of genes expressed in particular tissues, should help to elucidate gene functions.

Computational strategies will also contribute by annotating the draft with functional information. David Lipman, director of the National Center for Biotechnology Information in Bethesda, Maryland, says that algorithms developed by his centre and others are able to trawl the draft genome and find homologues to known genes "amazingly well". The forthcoming relaunch of the main



Collins: each new genomics initiative is valuable, but the challenge lies in juggling them all.

sequence database (see below) will let annotators navigate the genome with ease.

Francis Collins, director of the National Human Genome Research Institute, stresses that each of these initiatives is valuable. "Every full-length cDNA will add value," he says. "So will every mouse sequence read."

But at last week's meeting he warned of the problem of finding resources to support these initiatives without delaying the production of the finished sequence: "How we juggle all these things is challenging."

Some of Collins's advisers fear that the juggling will compromise the project's main

goal. "Somewhere in here we need to do something well," advisory council member Maynard Olson, of the University of Washington in Seattle, told the meeting.

Michael Morgan, chief executive of the Wellcome Trust's genome campus, which leads the British branch of the project, says the US's international partners are determined not to miss the 2003 target. "We must not lift our eyes from the goal of completing the genome," he says.

Originally, the Human Genome Project was intended to produce finished sequences, chromosome by chromosome. That resulted in the project's biggest milestone so far, the virtually complete sequence of chromosome 22 (see *Nature* **402**, 447; 1999). However, the project shifted strategies towards an interim goal of producing a draft of the entire genome after Celera Genomics, a sequencing company in Rockville, Maryland, joined the race.

Although the strategy has changed, the final goal should not, argues genome advisory council member Aravinda Chakravarti of Case Western University in Cleveland.

"It's not a question of either/or," says Robert Horvitz of the Massachusetts Institute of Technology, a member of the genome advisory council. "It's a question of balance." But that leaves Collins with a difficult balancing act.

Paul Smaglik

Revamped GenBank offers extra data links

Washington

GenBank, the leading human genome database, is about to undergo a facelift.

The National Center for Biotechnology Information (NCBI) in Bethesda, Maryland, will relaunch GenBank next month in a form that will be easier for users to navigate and will include a wealth of new 'annotation', such as information on genetic markers and gene and protein functions.

Users will be able to click on

part of an onscreen chromosome and retrieve information such as the predicted function of a partially sequenced gene.

Until now, this information has only been available for chromosome regions for which a contiguous DNA sequence is available. Much of this annotation will be generated automatically by computer algorithms.

Users will also be able to navigate more easily from various genetic markers in the database,

such as sequence tagged sites and expressed sequence tags. And they will find it easier to zoom in and out on specific chromosome regions, says David Lipman, director of the NCBI.

The new GenBank will also allow more links with the database of single nucleotide polymorphisms. This separate public database details the subtle variations in sequence found between individuals for certain genes.

P. S

GM debate must go global, says meeting...

Edinburgh

A permanent international forum should be set up to assess both the science and the social implications of genetically modified (GM) foods, according to key scientists at a meeting last week organized by the Organisation for Economic Co-operation and Development (OECD).

Leaders of the world's eight largest industrial nations (the G8 nations) are likely to be asked to endorse a proposal to create an appropriate panel.

The panel, proposed by Sir John Krebs, chairman of Britain's new Food Standards Agency, would be conceptually similar to the Intergovernmental Panel on Climate Change (IPCC), which looks at the science and implications of global warming.

Last year, the G8 heads of state asked the OECD to address the issue of health and scientific aspects of GM foods, after hearing demands for a moratorium on the commercial planting of GM crops before safety had been properly assessed.

"Such a mechanism, intended to inform politics in the future, must be firmly grounded in science," says Krebs, who chaired the Edinburgh meeting. "But it must go beyond the science. And it needs to be an international body because we are dealing with a global issue; whether or not consumers in one particular country reject



Emotive produce: protesters uproot GM plants.

GM foods, their development is already going ahead in several countries."

Sir Robert May, the UK government's chief scientific adviser, said that, just as the IPCC had been "very helpful" in the climate debate, something similar was needed on the GM issue.

"You need a mechanism for delineating the landscape, and for identifying things that we agree on and things that we do not agree on. An international panel would be a first step along this road," says May.

Participants on both sides of the debate have cautiously welcomed the proposal. Paul Muys, a spokesman for the European Association for BioIndustries, said his group was "inclined to see this in a positive way".

"Our approach is that everything should

be science-based," said Muys. "But there are other non-scientific issues that need to be dealt with as well."

Environmental groups present at the Edinburgh meeting were equally cautious, seeing potential value in an international panel able to assess the current state of scientific knowledge. But they were also concerned that it should not be science-driven, but start from a broader social agenda.

"Krebs has described such a panel as being 'science plus'," said Benedikt Haerlin, coordinator of Greenpeace International's Genetic Engineering Campaign. "But attempts to say scientists should decide which way to go has led us into just the situation that we now face. The forum should be based on a 'people plus science' approach; that is completely different."

If the idea is endorsed at the G8 summit in July, a key issue needing to be addressed is who should organize such an international panel. One leading candidate would be the OECD itself, which has been responsible for a series of studies and reports on the safety and regulatory aspects of biotechnology in recent years.

But such a prospect received a cool reaction from the environmental groups present. Peter Riley, for example, food campaigner for Friends of the Earth, expressed concern that the OECD — whose activities many still see as reflecting the views of the industrial community — has no tradition of giving non-governmental organizations a significant role in its deliberations.

Suman Sahi, president of the Gene Campaign in India, was equally sceptical. "We have had very bitter experiences with international forums," she said. "We have seen what happened with the [World Trade Organization]. If there is to be any internationalism on this issue, it has to come after discussion with the developing countries that have a view on the matter."

Another large question mark is the likely attitude of the United States. US officials are reluctant to make any public comment until the proposal has been developed further. But just as the IPCC has come under fire from US energy companies, many in the US life-science industry are thought to have reservations about the creation of any international panel that might legitimate tighter regulation of their activities.

One main point of agreement to emerge from the conference, however, according to a report presented to the final session by the meeting's rapporteurs (see box, left), was the need for "openness and transparency" in the policy process. Even the critics of GM food accepted that a global panel might be a step in this direction.

David Dickson

... and calls for openness and transparency

Edinburgh

The debate on genetically modified (GM) foods must become "more open, transparent and inclusive", concluded an international meeting in Edinburgh last week.

A draft report on the three-day conference, which was held to address the scientific and health aspects of GM foods (see main story), said it was agreed that the general public — as both consumers and citizens — "have valid points of view, which need to be understood and taken into account".

The draft also highlighted agreement on the fact that, although many people already eat GM foods, "no significant adverse effects have yet been detected on human health".

Areas of disagreement included whether GM foods in animal feed present a problem,

and also "who is to decide the use of [such] food for obtaining benefits as they are perceived in different parts of the world".

The conference, run jointly by the British government and the Organisation for Economic Co-operation and Development, was not intended to produce complete consensus among its 400 participants — which included representatives from government, industry and environmental groups — but rather to sketch out areas of agreement and disagreement.

It was generally agreed that, while concepts such as 'substantial equivalence' — the idea that a new GM food should be assessed partly on the basis of its similarity to known foodstuffs — are valuable tools for assessing risk, they need to be kept under review.

Sir John Krebs, who chaired

the conference, said that although substantial equivalence had been productively used for seven or eight years, "it is occasionally good to stand back and look at the picture as a whole".

Uncertainty remained on the potential long-term effects on human health and worker safety, as well as long-term environmental effects, including the impact on tropical zones.

The draft report also noted that, while there seemed to be agreement that risk assessment should be opened up and "should take into account non-positivist views", it was not obvious how this should be done in practice.

Krebs is to summarize the conference's main themes at a meeting of the heads of the eight largest industrialized nations in Tokyo in July. Details of the conference can be found on <http://www.oecd.org>

D.D.

Unesco 'worse than I imagined,' says new director

London

The new director of Unesco, the United Nations Educational, Scientific and Cultural Organisation, has given a damning report on the organization in his first diagnosis since taking office three months ago.

Addressing Unesco's executive board last month, Koichiro Matsuura said that the top-heaviness of Unesco's secretariat was "alarming", that there was a "severe mismatching of skills to actual needs", that the structure of the organization defied "all management principles and even logic". Staff morale, he said, was worryingly low. He added that existing management practices were often questionable, lines of authority were unclear and transparency was applied "very unevenly".

He also warned that Unesco was likely to overrun on staff costs by \$11 million over the next two years. He blamed the projected costs on the large number of appointments and promotions made by director Federico Mayor shortly before leaving the organization.

Only one of Mayor's 50 appointments and 55 reclassifications and personal promotions — mainly to director level — conformed to staff rules and regulations.

When he took over, Matsuura placed a freeze on these 'eleventh-hour promotions' and established a task force to examine each appointment. Many staff were offended by this. In January two members of staff went on a ten-day hunger strike in the entrance hall of Unesco's Paris headquarters (pictured, right).

Countries like the United States and Singapore, which are no longer Unesco members, will be watching events closely to see if Matsuura can, as he says, cut costs and improve the management of the organization. This could justify their re-entry.

The US State department called Matsuura's address significant: "We told Matsuura that it's important he create an atmosphere in which the rules are followed, the trains run on time, and there is clear transparency in personnel appointments and management practices". But at the same time, the US is not eager to see blame heaped on Matsuura's predecessor.

"We would not want this regrettable confusion and debate about eleventh-hour promotions to cloud the remarkable political turnabout Mayor brought about."



Desperate move: hunger striker Brinda Runghen.

Some commentators say Mayor is admired for his transformation of Unesco and his commitment to democratic values and freedom of information.

In 1993 the US General Accounting Office reviewed the management of Unesco and said that the organization had a culture of management reform. They found that significant progress had been made in the improving administration procedures.

John Fobes, chairman of Americans for the Universality of Unesco, says that Matsuura comes with only limited experience of Unesco through the World Heritage Program and will need time to find his way. Fobes' group has been reporting on the organisation since the demise of the US National Commission for Unesco in 1984.

Fobes says his group agrees that, even with the management problems facing Matsuura, it would be best for the US to work inside Unesco as soon as possible. He also hopes that additional specific project partnerships will be sought between US entities and Unesco.

In outlining his plans for reform, Matsuura remarked that "we have a house to live in and repair at the same time. Some of the repairs cannot wait. The leaking plumbing and faulty wiring has to be mended straight away."

"I must admit the real situation has turned out to be much worse than I imagined," added Matsuura, who also warned that the culture of Unesco as a whole needed to be changed.

Natasha Loder

China looks to west for economic growth

Beijing

Scientists are to play a leading role in a campaign launched by China to develop its economically disadvantaged western region. The region, covering an area of 5.4 million square kilometres, includes the Tibetan and Loess plateaux, and is home to many of China's poorest people. It also includes Shaanxi, which was the centre of ancient Chinese civilization in the Tang dynasty a thousand years ago.

The science ministry has been allocated 50 million yuan (US\$6.1 million) to spend in the western region over the coming year. It plans to build laboratories and engineering centres in local universities, to create an early warning system for monitoring ecological and environmental changes in key parts of the region, to set up an agricultural information network, and to conduct demonstration projects for boosting new technologies in environmental protection.

The Chinese Academy of Sciences will also invest 250 million yuan in its own development plan for the region to complement a training programme for PhD students it launched four years ago. Funds for that programme will be almost doubled.

The academy's new programme will study the mechanism of environmental deterioration in western China and establish pilot projects for sustainable development. The academy will also select practical technologies, such as water saving and clean energy technologies, to be used in the region, says Li Rui, deputy director of the academy's Institute of Water Conservation near Xi'an, the capital of Shaanxi.

The State Council, China's central government, has set up a special committee chaired by the prime minister, Zhu Rongji, to guide a government-wide campaign to build up infrastructure in western China. Funding will come from the issue of 100 billion yuan

in long-term treasury bonds, announced on 5 March.

But the committee faces a daunting task. Although Shaanxi has China's third highest density of universities and research institutes, young scientists prefer to move east, where salaries and working conditions are better.

"Strong in science does not necessarily mean strong in economy," says Li. "We need to find ways of applying scientific results, and using science to guide economic development." Technology transfer is particularly difficult in the western provinces, where industries are mostly state-run and tend to be burdened with debt, he says.

However, Li remains optimistic that some of the problems will be solved. "We will give incentives to encourage researchers from eastern institutes to take part in research projects in the west, and western researchers are also encouraged to conduct exchanges with their eastern colleagues."

Tian Xuewen

Expensive space crystal programme has produced little of scientific value, says panel

Washington

Space experiments in protein crystal growth have yielded no important results to date, says a report released last week by the US National Research Council (NRC).

But the space station's research facilities could be used to assess whether or not crystal growth in zero gravity could ever be worthwhile, says the report.

"At this time, one cannot point to a single case where a space-based crystallization effort was the crucial step in achieving a landmark scientific result," concluded a study panel chaired by Paul Sigler, who before his death in January was a professor of molecular biophysics and biochemistry at Yale University. "In many of the cases that have so far been listed as successful, the improvements obtained have been incremental rather than fundamental."

The report is just the latest criticism of the US space agency NASA's microgravity research programme, which has long touted protein crystal growth as a high-payoff area of research on board the space shuttle and eventually the space station. While the NRC was not so harsh in its judgement as a panel of the American Society for Cell Biology, which two years ago called for scrapping protein crystal growth experiments in space (see *Nature* **394**, 213; 1998), it left little doubt that past NASA claims of important research breakthroughs have been overhyped.

"To date, the impact of microgravity crystallization on structural biology as a whole has been extremely limited," says the NRC report. Sigler's committee found a number of reasons for this. Space-shuttle experiments have been brief and have suffered from the lack of a vibration-free environment. Only a small and insular group has been involved in the research so far. Although many private companies have signed on as partners for space experiments (which NASA press releases often emphasize), "not one has yet committed substantial financial resources".

Another problem is determining exactly what role microgravity has played in past successes. For example, when crystals of the restriction endonuclease *EcoRI* complexed with DNA (*EcoRI*-DNA) were grown in orbit, the resulting diffraction data were significantly better than those in similar samples grown on Earth. However, says the NRC panel, it was difficult to attribute the success to microgravity alone when advanced cryogenic techniques and synchrotron radiation analysis were also used.

In fact, the higher data resolutions that are now being achieved using synchrotron



Crystal gazers: predictions about the usefulness of space crystals have not proved true.

sources on the ground have wiped out one of the main advantages offered by space-grown samples — crystal size.

"Although the misconception that size is crucial may persist at NASA, scientists today are interested in crystallization methods that provide higher quality crystals," wrote the panel. And until the value of space-based crystals is proven in the case of specific research problems, the high cost of such research in orbit "is bound to engender resentment in the scientific community".

The NRC panel makes several recommendations. First, the space agency should take action to involve more people in space-based experiments, as the current recruitment process fails to attract the best scientists

and bioengineers. NASA might consider joint solicitations with the National Institutes of Health and the National Science Foundation.

The panel also advises NASA to instigate a high-profile grant programme to settle once and for all the usefulness of space-grown protein crystals. The space station, with its superior facilities and longer experiment runs, will be a suitable place to conduct such tests, it says.

"If none of the projects produces a space-grown crystal that enables a breakthrough for the structure determination of a biologically important macromolecular assembly, then NASA should be prepared to terminate its protein crystal growth program," the report concludes.

Tony Reichhardt

Geneticists oppose consent ruling

Washington

The chief of the National Human Genome Research Institute has warned against a move to require human geneticists to obtain informed consent from family members, as well as primary subjects, when they solicit people's family histories.

Francis Collins has challenged a recent interpretation of government rules by the Office for Protection from Research Risks (OPRR). It said that when researchers obtain private, identifiable information

about individuals — whether primary subjects or their family members — those individuals become by definition human subjects, and their informed consent must therefore be sought for participation.

According to Collins, this "could render a large proportion of current research in human genetics impracticable — with highly detrimental consequences to ultimate public benefit".

The case in question involves a twin study at Virginia Commonwealth

► University, where the OPRR shut down federally funded research in December, citing numerous deficiencies in human subjects' protection. (The research is now being resumed.)

OPRR's investigation at the university was prompted by a complaint from Richard Curtin, a budget analyst at the US Department of Defense who is the father of college-age twins. One of them received a mailed questionnaire from Linda Corey, a professor of human genetics at the university, that solicited information for the Mid-Atlantic Twin Registry, a 20-year-old registry of 30,000 twin pairs used by medical researchers. The questionnaire asked about the occurrence of hundreds of medical conditions, including abnormal genitalia, alcoholism and infertility, in the twin and family members.

It was "a total invasion" of privacy, says Curtin, who adds that his security clearance at the Department of Defense could be revoked if he suffered from conditions such as mental illness.

"It appears that the [university ethics board] failed to consider the potential social, psychological, and legal risks" presented to twin subjects' family members by the collection of their detailed medical and social information without consent, OPRR wrote to university officials.

But in a letter in January, Collins told OPRR director Gary Ellis that he considered the presumption of risk to family members was unreasonable. "The OPRR's [position] represents a new policy that does not appear to have been informed by broad scientific or public input," he charged, adding that he has "deep concerns" about the decision.

Ellis, in a response to Collins on 22 February, said that OPRR had simply applied existing human subject protection regulations to a particular case, and not implemented a new policy or imposed new general rules.

"Please do not infer any general rule-making by OPRR regarding informed consent from family members beyond the specifics of this particular research activity," said Ellis.

The American Society of Human Genetics (ASHG) and the National Institute for Child Health and Human Development have both contacted Ellis on the matter.

"We are very concerned about the fact that the collection of family history, which is a very important part of human genetics research, could be inhibited by too much control," says Uta Francke, the previous president of ASHG and a professor of genetics at Stanford University.

Meredith Wadman

Panel will seek 'appropriate' AIDS goals for South Africa

Cape Town

South Africa's health department is setting up a panel of experts to tackle the AIDS epidemic, health minister Manto Tshabalala-Msimang confirmed this week.

The panel of some 30 local and international experts will "explore all aspects of... developing prevention and treatment strategies that are appropriate to the African reality," she says in a press release.

The minister is hoping to reassure AIDS activists, who have accused the government of wilfully mismanaging the epidemic in South Africa, that the panel will be free to work to its own conclusions.

But she has not confirmed or denied the rumour that controversial biochemist Peter Duesberg of the University of California at Berkeley — who claims that the HIV virus is not the cause of AIDS — may be on the panel.

"Those with more extreme views are unlikely to participate because we are looking for a consensus," she says. AIDS activists continue to suspect that the government line supports the Duesberg claim.

The panel will review the general prevention and treatment (as well as the causes and diagnosis) of HIV/AIDS and opportunistic infections. It will also review the prevention of infection following rape or needle-stick injuries, and from mother to child.

The South African government recently refused to supply anti-retroviral drugs such as AZT to pregnant women within the state health system. It apparently believes that the



Tshabalala-Msimang: 'no hidden agendas'.

risks of using the drugs outweigh the benefits, despite advice to the contrary from different expert groups (see *Nature* 403, 692; 2000). Tshabalala-Msimang says this decision could be reconsidered if the panel convincingly shows that treatment would be effective. But an "ingenious solution" to the difficulties of financing the treatment would need to be found in such a case, she says.

Tshabalala-Msimang denies claims by the Treatment Action Campaign (TAC) that the panel is "a justification for the immoral, unscientific and unlawful decision" not to give the drugs to pregnant women. "I hope the work of the panel will demonstrate that we have no hidden agendas," she says.

The TAC has challenged the minister and her advisers to publish evidence from any scientific study to prove that provision of AZT is not economically feasible in South Africa, or that AZT is toxic to mother or child when given to women in the last trimester of pregnancy.

Michael Cherry

\$350m gift boosts MIT brain power

Boston

The Massachusetts Institute of Technology has a bold plan to make itself a world leader in neuroscience. Last week, MIT announced the creation of an ambitious institute to study the human brain.

It will be the centrepiece of a new neuroscience complex, scheduled for completion in 2004. The McGovern Institute for Brain Research will be joined in the complex by an expanded version of MIT's Center for Learning and Memory (CLM) and a \$20 million centre for brain imaging. "With all the resources here, MIT should stand among the best in this field," says MIT molecular biologist Phillip Sharp, who will direct the institute.

Patrick McGovern, founder of computer publishing giant the International Data Group, and his wife Lore Harp McGovern, a

high-tech entrepreneur, will give \$350 million over 20 years — the largest gift ever pledged to a US university. They picked MIT, from which Patrick McGovern graduated in 1959, because of its reputation for fostering interdisciplinary research.

The McGovern Institute will house 16 labs and 300 staff. Its model is the Whitehead Institute for Biomedical Research, also at MIT, where Lore Harp McGovern chairs the Board of Associates and her husband is a trustee. "I was impressed with how much progress they've made, for example, in understanding the causes of cancer and other diseases," says Patrick McGovern. "It shows what can be done at a mission-oriented centre if you bring the right people together."

For Sharp, who won the Nobel prize in medicine in 1993 for his work on RNA



Think about it: MRI scan of a healthy brain.

splicing, the move into neuroscience is a shift in direction. "This field is ready for rapid advance," he says. Brain imaging techniques, tools from molecular biology and genetics, insights from the Human Genome Project, and increasing computer power offer huge potential, he says. "The challenge now is to put together a first-rate programme." He expects the centre to be fully developed within about 10 years.

Sharp is working with Tomaso Poggio, an MIT computational neuroscientist, to chart future research directions. "We will have several years, if not longer, to make decisions about the range of programmes," Sharp says.

The complex will also include the \$20 million Martinos Imaging Center and the CLM. The imaging centre, to be shared with Harvard University and Harvard Medical School, will work on technologies such as magnetic resonance imaging (MRI), positron emission tomography, and optical imaging. "The idea is to use our technical prowess to advance these tools even further," says Mriganka Sur, who chairs MIT's Department of Brain and Cognitive Sciences.

The CLM, headed by MIT Nobel laureate Susumu Tonegawa, has seven investigators, with five more staff to be hired. It receives about half its funding, US\$3.5 million per year, from the RIKEN Institute, a Japanese government agency — a figure expected to grow to US\$5 million. Tonegawa predicts that the CLM and the McGovern Institute, working side by side, "will become a major force in neuroscience".

The trick now is figuring out how to make the various pieces fit together. The tentative plan is for the CLM, which investigates learning, memory and neuroplasticity, to continue its emphasis on molecular- and cellular-level processes, while the McGovern will focus on higher-level functions, such as information processing in the brain.

"We want enough overlap so we can interact without duplicating each other's efforts," says Sharp. "It's easy to identify broad themes of what we'd like to see emerge. But we won't know the specifics until we do the work."

Steve Nadis

Indian research budget favours defence

New Delhi

The Indian government has kept its promise to significantly boost the national research budget — but the main beneficiary will be the military.

The recent clash with Pakistan in the Kargil area of the Himalayas, which exposed India's weakness in satellite surveillance and mountain warfare, lies behind the preferential funding increases for military-related research.

Military research now absorbs more than half of the research budget, which, at 120,654 million rupees (US\$2.8 billion), is nearly 20 per cent up on last year (see table).

A large chunk of the additional funding will be used to develop and build five remote-sensing satellites, three of which will be high resolution. Although the government claims that these satellites are intended for both military and civilian applications, sources within the science ministry say their main application will be military.

The defence services had been demanding dedicated satellites for surveillance for some time, says Uday Bhaskar, deputy director of the Institute for Defence Studies and Analysis in New Delhi, adding that the Kargil conflict seems to have clinched it.

The Department of Atomic Energy intends to use part of its additional 2,270 million rupees to develop intense electron-beam machines that can potentially knock out enemy missiles.

It will also use some of the new funds to develop a 500-MW prototype fast breeder reactor which will use plutonium produced by Indian power reactors.

Science secretary Valangiman Ramamurthi says that, although more military research is needed for national security, the scientific community in general is "quite happy" with the budget. "Allocations for almost all science departments have gone up



Going great guns: the Indian militia in Kargil.

between 14 and 20 per cent compared to last year," he says.

"Basic research did not get as much as we wanted," he admits, but the government has allocated additional funds for the two new research centres — the National Brain Research Centre (500 million rupees) and the National Centre for Plant Genome Research (350 million rupees), both in New Delhi.

Ramamurthi is particularly happy that 500 million rupees has been earmarked for some of the technology projects identified in the Vision-2020 document prepared in 1997 by the Technology Information Forecasting and Assessment Council, a government-industry think-tank. "The money for this has come at the right time," he says. Projects such as the use of fly ash (a waste product from coal-based power stations) as building material and the development of hydrogen as an energy source will now be taken up seriously, he says.

The science budget also provides 500 million rupees for launching the New Millennium Indian Technology Leadership Initiative. Ragunath Mashelkar, secretary to the Department of Scientific and Industrial Research, says this initiative will "focus on four or five areas which would fulfil the national objectives of a global leadership in technology". The government, industry and venture capitalists will team up to accomplish this, he says.

India's 200 patent offices will receive 750 million rupees for modernization. Scientists have particularly welcomed the announcement that universities and research institutes can keep all the royalties earned from patents, amending existing rules requiring patent holders to share royalties with agencies that funded their research. **K. S. Jayaraman**

Table: Indian science budget (in million rupees)

	1999	2000
Defence	20,954	22,742
Space	17,259	20,192
Atomic energy	13,791	16,079
Agriculture	13,040	14,046
Industrial	8,237	9,704
Environment	7,180	9,650
Dept of S&T	6,173	7,798
Medical	7,519	9,139
Non-conventional energy	3,195	4,443
Information technology	1,950	3,920
Oceanography	1,067	1,580
Biotechnology	1,282	1,361
Total	101,652	120,654

Rival demands sink genome alliance plans

London

Secret talks between rival teams racing to complete the sequencing of the human genome have failed, dashing hopes of collaboration and leaving a trail of recriminations.

The two teams — the publicly funded Human Genome Project (HGP) and the private company Celera — had been discussing forming an alliance to pool their resources, but, as in a previous attempt to collaborate (see *Nature* 397, 93; 1999), they stumbled primarily over the issue of public access to data.

Although the HGP is opposed in principle to any kind of privileged access to data, it had offered a compromise allowing Celera some rights to any data generated as a collaboration for twelve months, in order to speed up completion of the sequence.

Celera's compromise was that the consensus genome could be released publicly at the time of completion, but only on the condition that the data could not be used to compete with Celera's position as a database provider for three to five years. All researchers would have unrestricted use of primary data, says a Celera spokesman, but data merged from the HGP and Celera projects and assembled by Celera "should not be given to those competing with us in the database business for a period of time".

Both sides are now bristling with anger. In frustration, the UK's Wellcome Trust, a major participant in the HGP, released last weekend a confidential letter sent to Celera by the HGP, numbering the issues it says were

hampering talks and asking for a response before 6 March 2000.

The letter — whose signatories include Francis Collins, director of the US National Human Genome Research Institute and Harold Varmus, director of the US National Institutes of Health — outlines the HGP's concerns about Celera's desire to have its commercial rights extended to research applications. These include the construction of genome chips, large primer sets, or applications to proteomics and analysis of regulatory sequences.

"Our position is that we would do nothing in collaboration that would affect our ability to develop the intellectual property on our discoveries," says Paul Gilman, director of policy planning at Celera.



The letter says that the HGP is troubled by what it perceives as Celera's intention to publish under its own name merged data in a peer-reviewed journal. "We would give appropriate acknowledgement or credit for [HGP scientists'] contributions — if they chose not to collaborate," says Gilman.

Michael Morgan, chief executive of the Wellcome Trust's genome campus, says he felt forced to release the letter in advance of the response deadline because the Trust had been "frustrated" by Celera's lack of response. "Negotiations seemed to be going nowhere," he says. At the same time, Celera was making optimistic statements in public, expressing hope that a collaboration would get off the ground, says Morgan.

Celera counters that the early release of the letter was neither ethical nor "in the spirit of co-operation". This week Venter told *The Washington Post* that releasing the letter was "a low-life thing to do". Meanwhile, John Sulston, director of the Sanger sequencing centre, told the *BBC Today* programme that Celera's taking public data and selling it with their own amounted to something of a "con-job".

Hopes are now dwindling that the groups could work together. Morgan says this is not the end of the idea but warns that the possibility for collaboration is dwindling. "The train is rapidly leaving the station," he says.

Gilman declined to say where the collaboration stands "until we have heard directly from the National Institutes of Health or the Wellcome Trust".

Natasha Loder

Biology back issues free as publishers walk HighWire

Paris

Eighty-three life sciences journals will make their back issues — representing more than 137,000 articles — free on the web through High Wire Press.

HighWire Press is a not-for-profit outfit set up in 1995 by Stanford University Libraries and Academic Information Resources to help universities and societies to publish on the web at low cost.

This new agreement with the 83 journals makes HighWire the world's second-largest scientific repository, after the US space agency NASA's Astrophysics Data System. It already dwarfs the nascent PubMed Central (PMC) initiative that has been set up by the US National Institutes of Health — whose goal is to create a free global website for the entire life sciences literature (see *Nature* 401, 6 & 626; 1999).

PMC went live last month, but has so far attracted only a handful of publishers, many of whose back issues are also available from

HighWire, such as *Molecular Biology of the Cell* and *Proceedings of the National Academy of Sciences* (PNAS).

Most journals taking part in the HighWire scheme will make their content free one or two years after print publication (see <http://highwire.stanford.edu/lists/freeart.dtl>) and frequently also on a trial basis. But PNAS is making its content freely available one month after print publication, and *Molecular Biology of the Cell* after two.

David Lipman, director of the National Center for Biotechnology Information at the NIH and one of the main architects of PMC, puts a brave face on this situation.

The HighWire initiative is "a great thing", he says. "It shows in principle that journals are willing to make back content free and therefore should be willing to take part in PMC."

"Our goal is to provide free access for all life sciences. HighWire can only do it for journals that are paying them."

But some editors are already questioning the need for PMC. Ira Mellman, editor-in-chief of *The Journal of Cell Biology*, which is making its back issues free on HighWire after 18 months, says: "Allowing journals and publishers to select their own host sites makes more sense than trying to enforce a common solution on all."

Another journal editor adds: "Virtually everything they have proposed in this regard has been developed by HighWire and other commercial sites. Free access is already provided to literally anyone at an academic or commercial institution."

HighWire's publisher, Michael Keller, says he intends "to expand the coverage of our disciplines".

Momentum for making past content freely available is growing, Keller claims. "The success of the free back issues effort will be so considerable that many publishers will want to go in that direction to achieve that social goal," he says.

Declan Butler

Russian troops free Polish ecologists from Chechnya

Moscow Two Polish ecologists kidnapped last August by Chechen bandits have been rescued by Russian forces. Zofia Fiszer-Malanowska of the Warsaw Ecology Institute of the Polish Academy of Sciences and Eva Marchwinska-Wyrwal of the Ecology Institute in Katowice were freed in the Argun ravine, Chechnya.

After capturing the scientists in the neighbouring republic of Dagestan, the kidnappers demanded a ransom of \$1 million for each prisoner. However, the Polish authorities refused to pay (see *Nature* **402**, 114; 1999).

Details of the operation that freed the two scientists remain secret. Both are now back in Warsaw.

Texaco pulls out of US climate lobby group

Washington The Global Climate Coalition, a Washington-based lobby group that opposes legislation to curtail global warming, has lost three members in the past four months. Last week, Texaco notified the coalition that it would not renew its membership for 2000. Daimler-Chrysler withdrew in January, and

Ford said it would not renew its membership in December.

Glenn Kelly, the coalition's executive director, is putting a brave face on the defections. He claims that the three companies still support the coalition's opposition to the Kyoto Protocol. The departures will not affect the coalition's lobbying leverage at all, Kelly asserts.

German medics oppose embryo laws

Munich The German Chamber of Physicians has issued a proposal calling for a relaxation of rules that forbid preimplantation genetic diagnosis.

Germany's strict embryo-protection law currently bans preimplantation diagnosis on ethical grounds. But the Chamber of Physicians argues that there is an increase in "medical tourism" by German couples seeking advice in neighbouring countries. Preimplantation diagnosis is allowed in most countries of the European Union. Worldwide, 400 couples have used the technique, and more than 100 babies have been born following preimplantation diagnosis.

The Chamber of Physicians would like preimplantation diagnosis to be legal in Germany for some medical conditions, such as monogenetic diseases or parental

chromosome defects. It says that such diagnosis should be limited to genes associated with diseases for which the parents have a genetic predisposition.

The German health ministry has indicated that it may be prepared in the future to modify the embryo protection law — but not yet. Opponents of preimplantation diagnosis, such as the Berlin-based Gene Ethical Network, fear that it could foster eugenics.

Centre for endangered species boosts research

San Diego The Center for Reproduction of Endangered Species at the San Diego Zoological Society, known worldwide for its condor and panda projects, is undergoing a major research expansion.

The centre is seeking 12 postdoctoral researchers for international studies on endangered species and is building a \$20 million laboratory facility. It has also hired a new director, Alan Dixon, who left the University of Cambridge last autumn.

Research projects for the new postdocs will be funded either through the zoological society — which has an annual operating budget of \$100 million for its facilities, including two zoos — or through grants secured competitively from funding agencies, officials said.

Compton observatory faces reprieve

Washington The US space agency NASA's Compton Gamma Ray Observatory will not be deliberately destroyed until June at the earliest, and may remain in orbit for years if engineers can persuade nervous NASA managers that a new technique for controlling the spacecraft without gyroscopes will work. After one of the observatory's on-board gyroscopes failed in December, NASA started making plans to bring the spacecraft down in a fiery re-entry this month to avoid the risk of an uncontrolled crash over a populated area (see *Nature* 403, 232; 2000).

The plan now calls for a June de-orbiting, says project scientist Neil Gehrels of the Goddard Space Flight Center. But he is hopeful that the new control technique, which has proven feasible in simulations, will persuade NASA to leave Compton in space. A decision is expected on around 23 March.

NIH finds more adenovirus 'serious adverse events'

Washington The National Institutes of Health (NIH) has revised upwards by 50 per cent the number of serious adverse events that investigators failed to report promptly in gene therapy trials using adenoviral vectors.

Adenoviruses were used in the trial in

which an 18-year-old volunteer died last September at the University of Pennsylvania (see *Nature* 401, 517; 1999). Following that tragedy, the NIH called for more information on adverse reactions. By November, 652 events had been reported. Last week, Lana Skirboll, director of NIH's Office of Science Policy, testified to the National Bioethics Advisory Commission that 970 serious adverse events have now been reported in about 70 trials over seven years.

Partnership reveals malaria projects

Paris The Medicines for Malaria Venture, a public/private-sector collaboration which aims to boost research into malaria drugs, has selected its first three projects. Although details of the projects have not yet been released, *Nature* has learnt that the winners are: a project involving Glaxo Wellcome, the London School of Hygiene and Tropical Medicine, and the University of Bristol; one involving Hoffmann La Roche, the University of Nebraska, and the Swiss Tropical Institute; and one by SmithKline Beecham and the University of California at San Francisco.

All three multimillion-dollar projects are thought to be investigations into new families of drug targets that are less likely than current ones to succumb to parasite resistance.

Japanese universities may keep half of royalties

Tokyo Public universities could be authorized to withhold half of the royalties on patented inventions, the Japanese Ministry of Education has announced. At present, royalties are paid directly into the national treasuries but the ministry hopes to provide incentives for entrepreneurial activities within academia.

Observers have welcomed the new scheme but point out that many universities lack the resources to manage intellectual property rights. Although the number of papers co-authored with industry has increased substantially in recent years, the number of patents held by Japanese universities is insignificant and has stagnated.

Drug companies agree to donate vaccines

Washington Some of the world's leading drug companies are to donate millions of doses of vaccines to combat diseases such as malaria and polio in developing countries.

The donations, valued at more than \$150 million, follow a call by the US president, Bill Clinton, for an international endorsement of his Millennium Vaccine Initiative. It would mean a major boost in vaccine efforts and a \$1 billion tax credit for drug companies that invest in vaccine development.

Why private institutions alone will not do enough to protect biodiversity

Sir—The recent interest by private-sector institutions in conserving biodiversity is encouraging^{1–3}, but in our view government investment remains essential.

In the first place, the private sector is not well equipped to provide public goods related to the global environment. Once provided, public goods are freely available. Ecosystem resilience, aesthetic and existence values, and the option values of biological resources all contain significant public goods related to environmental security⁴. But the respective roles of public and private sectors in providing biodiversity are not straightforward, because most ecological units generate both public and private goods. Private investment, such as in a nature reserve, often creates public goods, but this cannot always be expected. For example, it may not adequately conserve non-economic (or unappealing) species, nor preserve habitats at the expense of profitable development alternatives.

Also, the rate of biodiversity decline demands action now. Private conservation efforts are far from adequate, especially in developing countries. Delaying widespread action until private-sector investment has greatly expanded would be hazardous.

We think that government intervention to promote biodiversity conservation is essential in three respects. First, governments must provide most of the increased funding needed to conserve a representative sample of the Earth's ecosystems, at least until such requirements can be met by the private sector. We estimate that this might cost around US\$27.5 billion annually, compared to the \$6 billion currently spent by governments, private sector and foreign donor institutions⁵. If the international community values the global public goods provided by biodiversity, it must greatly increase its financing of conservation, particularly in developing countries.

Second, private-sector conservation rarely succeeds without government incentives, either in developed⁶ or developing⁷ countries. These include expanding legal definitions of property rights to cover environmental resources; reducing costs through help with information technology, administration and enforcement; and providing an effective framework of regulation.

Third, governments must cut environmentally perverse subsidies in natural-resource sectors⁸. These not only harm biodiversity directly but also inflate the costs of conservation by increasing the profitability of non-sustainable options.

Anyone who doubts that governmental

and intergovernmental policy can profoundly alter biodiversity need only examine the impact of schemes such as the EU's Common Agricultural Policy. By subsidizing the over-production of food, this has greatly accelerated agricultural intensification, resulting in the widespread and rapid decline of many birds, insects, plants and traditional farming landscapes⁹. Reducing such subsidies would save taxpayers' money, free government funds for conservation, and reduce the costs to public and private sectors of expanding efforts to stem the loss of biodiversity.

Alexander James*, **Kevin J. Gaston†**,
Andrew Balmford‡

*Department of Land Economy, University of Cambridge, Cambridge CB3 9EP, UK

†Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

‡Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

1. Chapela, I. *Nature* **403**, 129 (2000).
2. Brown, M. & Leal, D. *Nature* **403**, 129 (2000).
3. Daily, G. & Walker, B. *Nature* **403**, 243–245 (2000).
4. Daily, G. (ed.) *Nature's Services* (Island, Washington, 1997).
5. James, A. *et al.* *Nature* **401**, 323–324 (1999).
6. Hodge, I. in *Changing Rural Policy in Britain* (eds Curry, N. & Owen, S.) 127–144 (Countryside and Community Press, Cheltenham, 1996).
7. Attwell, C. A. M. *Oryx* **34**, 3–5 (2000).
8. Myers, N. *Nature* **392**, 327–328 (1998).
9. Krebs, J. R. *et al.* *Nature* **400**, 611–612 (1999).

Alzheimer's research is vital in work on ageing

Sir—Leonard Hayflick makes some good points about ageing and ageing research in his Millennium Essay (*Nature* **403**, 365; 2000). I agree that “humans, and the pet and zoo animals that we choose to protect, are the only species in which large numbers experience ageing”. There has been much conjecture regarding the functions of post-reproductive life in humans and the lack of post-reproductive life in other primates.

However, I think Hayflick's comments about funding by the US National Institute on Aging for Alzheimer's research are far off the mark. Although it is true that humans have a life-expectancy far beyond that for which our evolutionary history has prepared us, it does not follow that our necessary goal is the extension of our life span even further. Far more important is the goal of enhancing the quality of life within the lifespan we already have.

Even modest advances in delaying the onset of Alzheimer's and other age-related diseases can have a tremendous personal and economic impact. Fundamental research on the biology of ageing is immensely important. There is mutually beneficial interplay between research aimed at eradicating Alzheimer's and that directed at understanding ageing in cells,

worms, flies, mice, monkeys, apes and humans. Reducing funding for Alzheimer's research is one of the least satisfactory ways to increase funding for the basic scientific study of ageing.

Joseph M. Erwin

Division of Neurobiology and Behavior, Bioqual Inc., 9600 Medical Center Drive, Rockville, Maryland 20850, USA

Opportunism knocks?

Sir—Bernd Legler and Guy Moore¹ comment in Correspondence on the evaluation of East German institutes by the Wissenschaftsrat and the intellectual and moral qualities of the scientists examined.

First, we are not aware of interviews held in English during evaluation of institutions in the former East German Academy of Sciences (a statement² attributed to Jens Reich). Further, we found the evaluation here at the Institute of Plant Genetics and Crop Plant Research at Gatersleben to be a fair process.

Second, though East German academics in general spoke English less fluently than western colleagues, most researchers in Academy of Sciences institutes were able to give lectures in English; only a few would have done better in Russian. Scientific matters were discussed in English for preference, even with Russian colleagues. At East German secondary schools, English was usually the second (compulsory) foreign language, and the same was true for faculties of natural sciences at universities.

Third, Legler and Moore describe the (few) East German heads of institutions and departments who gained such positions after the fall of the Wall as opportunists “biting the hand that had once fed them”. The truth is that most of these scientists occupied third-rank positions during the final period of East Germany's existence because they stood by their principles. Although well qualified, they had no chance of promotion in a system that rewarded obedience rather than scientific merit.

The scientists themselves had different selection criteria from those of the political ruling caste. At Gatersleben, for instance, a scientific council was elected by the entire scientific staff as early as 1989. This group pushed through the replacement of the former department heads and recruited a new board of directors that was confirmed by the Wissenschaftsrat in the course of its post-unification evaluation of the institute.

Maybe there have been individual cases of opportunism—but even so, it is certainly not an East German invention.

Ingo Schubert, Ulrich Wobus

Institute of Plant Genetics and Crop Plant Research, D-06466 Gatersleben, Germany

1. Legler, B. & Moore, G. *Nature* **403**, 130 (2000).
2. *Nature* **401**, 637 (1999).

Rape as an adaptation

Is this contentious hypothesis advocacy, not science?

A Natural History of Rape: Biological Bases of Sexual Coercion

by Randy Thornhill & Craig T. Palmer
MIT Press: 2000. 272 pp. \$28.95, £17.95

Jerry A. Coyne and Andrew Berry

In *A Natural History of Rape*, Randy Thornhill and Craig Palmer argue that rape is an adaptation — that it has evolved to increase the reproductive success of men who would otherwise have little sexual access to women. Their analysis of rape then forms the basis of a protracted sales pitch for evolutionary psychology, the latest incarnation of sociobiology: not only do the authors believe that this should be the explanatory model of choice in the human behavioural sciences, but they also want to see its insights incorporated into social policy. Thus, in a single, slim volume, Thornhill and Palmer give us both an inflammatory analysis of a sensitive topic, and a manifesto outlining evolutionary biology's future conquest of the social sciences.

In the furore that has greeted the book's publication, the scientific evidence for the authors' arguments has been largely ignored. However, it is here that we must start. If their specific claims about rape are not scientifically sound, then the authors' grand vision of the centrality of natural selection to every aspect of our behaviour collapses as well. In their media appearances, Thornhill and Palmer cloak themselves in the authority of science, implying that the controversy over their ideas is purely political, and that the underlying biology is unimpeachable. This is a serious misrepresentation.

What persuasiveness the book does possess rests on an ingenious rhetorical trick. The authors lay out two alternative evolutionary hypotheses: rape is either a 'specific adaptation' (that is, natural selection explicitly promoted the act) or a 'by-product of evolution' (there was no direct selection for rape; rather it is an accidental product of selection for, say, male promiscuity and aggression). Readers unconvinced by the specific-adaptation argument therefore find themselves embracing by default the by-product alternative. Either way, Thornhill and Palmer claim a convert.

But what, in behavioural terms, is an evolutionary by-product? Everything that is not a specific adaptation. Thus, playing the piano — an activity unlikely to have been instrumental in the evolution of the brain — is an evolutionary by-product, because it depends on a brain that was itself produced by natural selection. If every human behaviour can be seen as a by-product of evolution,

then the by-product idea is useless, for a theory that explains everything is merely a truism. The claims that rape and playing the piano are by-products of evolution are claims without content.

It is not surprising, then, that *A Natural History of Rape* is largely an argument for the specific-adaptation theory. The authors' evidence, however, either fails to support their case, is presented in a misleading and/or biased way, or equally supports alternative explanations. The following examples illustrate each of these failings.

First, Thornhill and Palmer make much of the claim that rape victims tend to be in their prime reproductive years, suggesting that reproduction is indeed a central part of the rapist's agenda. But the data they present contradict this claim. In a 1992 survey that attempted to deal with the substantial statistical problem of unreported rape, 29% of US rape victims were under the age of 11. As that age group comprises approximately 15% of the female population, under-11s were *over-represented* among rape victims by a factor of two. So invested are the authors in their specific-adaptation hypothesis that they try

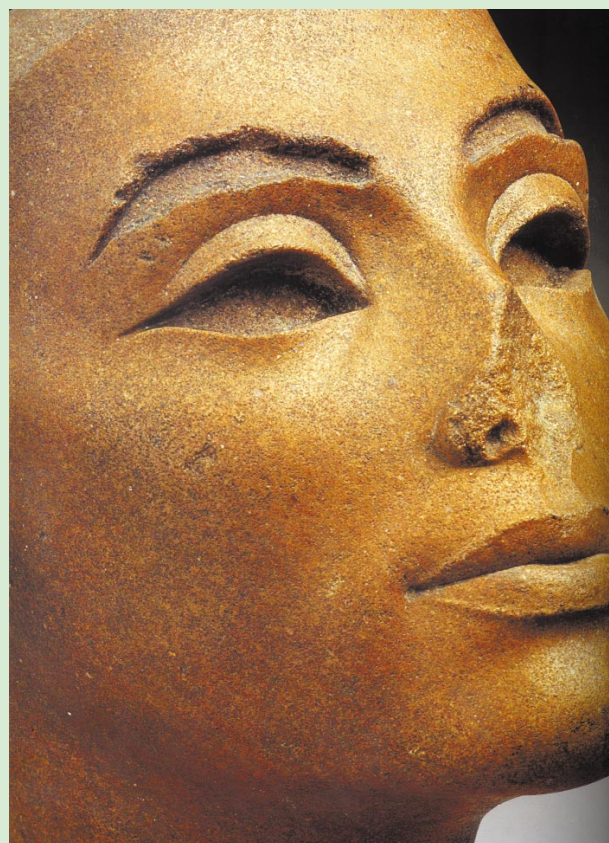
to explain this non-adaptive anomaly by noting that the data do not indicate the "proportion of the victims under 11 who were exhibiting secondary sexual traits". Further, "the increasingly early age of menarche in Western females contributes to the enhanced sexual attractiveness of some females under 12". In the end, the hopelessness of this special pleading merely draws attention to the failure of the data to support the authors' hypothesis.

Second, the authors contend that, based on sociological studies, female rape victims of reproductive age are more traumatized by the experience than are women either too old or too young to reproduce. The rationale is that reproductive-age women are in effect mourning the lost opportunity for mate choice which rape, in the world view of evolutionary psychology, represents to them. The authors see this apparent heterogeneity in the reaction to rape as supporting their claims about the reproductive essence of the act.

In checking the cited reference (one of its authors is Thornhill himself), we find that the original work's conclusions differ

A celebration of Egyptian history

This statue of Queen Nefertiti, wife of Akhenaten (1353–1335 BC), was discovered below the foundations of the palace of Merneptah at Memphis in Egypt. It now rests in the Egyptian Museum in Cairo as part of the world's finest and most celebrated collection of Egyptian artefacts. The collection and its history are described in *The Cairo Museum Masterpieces of Egyptian Art* (Thames & Hudson, £45), edited by Francesco Tiradritti, with abundant, splendid photographs by Araldo de Luca.



book reviews

critically from those given in the book. According to Thornhill and Palmer, the cited study shows post-rape trauma to be higher in reproductive-age women (age 12–44) than in the two other age classes (under 12 and over 44). In fact, the data show that the only heterogeneity in response to rape comes from the under-12 class: the over-44 class is just as traumatized as the 12–44 one. However, when the over-44 and under-12 classes are pooled, the under-12 effect of less trauma makes this combined 'non-reproductive' class significantly different from the 12–44 one. The authors have used statistical sleight of hand to buttress their argument. And we need hardly point out that the relative lack of trauma in the youngest age group may be unrelated to sexual immaturity: rather, children may be less able to express their feelings. Furthermore, the original study's data are questionable because much of the assessment of trauma in the under-12 class was necessarily based on reports from the child's care-givers rather than from the child herself. Direct comparison of observer-reported and self-reported data on such a subjective issue is extremely problematic.

Finally, the fact that women of reproductive age experience more violence during rape than do older women or children — suggesting that they fight back harder — is taken by the authors as evidence that they have more to defend. There is, they contend, more at stake — reproduction, no less — for reproductive-age women. While it is true that women of reproductive age who resist

rape may be partly motivated by the fear of unwanted conception, it is also true that such women, at the peak of their bodily strength, are most physically capable of fighting back. Children cannot fight off a full-grown man, and older women may also find resistance beyond them. In exclusively championing their preferred explanation of a phenomenon, even when it is less plausible than alternatives, the authors reveal their true colours. *A Natural History of Rape* is advocacy, not science.

We have highlighted just three examples of the book's flawed arguments. There are many more. The evidence that rape is a specific adaptation is weak at best. In keeping with the traditions established early in the evolution of sociobiology, Thornhill and Palmer's evidence comes down to a series of untestable 'just-so' stories.

Sociobiological approaches to human behaviour may yield interesting insights. But it is disciplinary hubris — a long-standing feature of evolutionary psychology — to suppose that natural selection underlies our every action. Because of the central role of reproduction in Darwin's theory, sexual behaviour is, in principle, a good candidate for fruitful sociobiological study, but even here it usually fails dismally. The most imaginative and committed sociobiologist would be hard-pressed to show that masturbation, sadomasochism, bestiality, and pornography's enthusiasm for high heels are all direct adaptations. In its insistence on forcing everything into the strait-jacket of adaptation, evolutionary psychology offers a

woefully incomplete perspective on human behaviour. Thornhill and Palmer have inadvertently revealed just how deficient that perspective is. ■

Jerry A. Coyne is in the Department of Ecology and Evolution, University of Chicago, 1101 East 57 Street, Chicago, Illinois 60637, USA.

Andrew Berry is at the Museum of Comparative Zoology Labs, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.

Amassing the case for the defence

Evolutionary Catastrophes: The Science of Mass Extinctions

by Vincent Courtillot

Cambridge University Press: 1999. 173 pp. £14.95

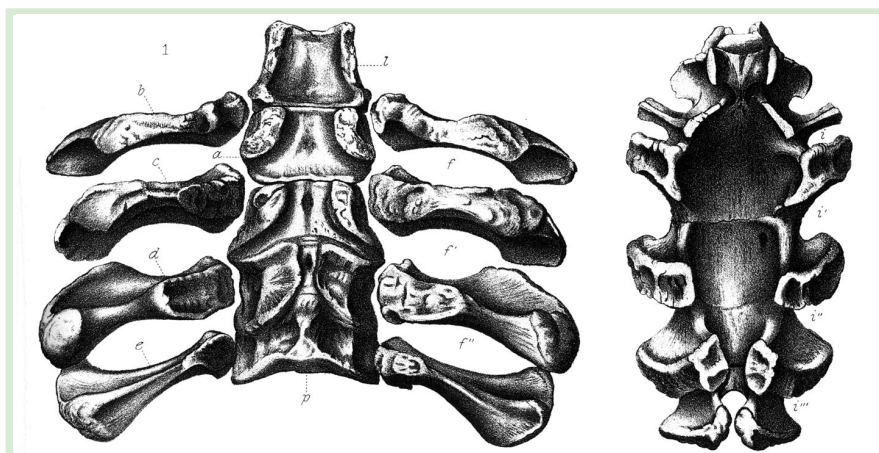
Douglas Palmer

Scientists who are not geologically minded may be puzzled by the apparent inability of geologists to agree on whether the dinosaurs, ammonites and their ilk were or were not 'polished off' 65 million years ago by the impact of an extraterrestrial bolide. Some geologists, such as Walter Alvarez and his father, Nobel prizewinning physicist Luis Alvarez, have argued vehemently for an impact event in the area of today's Yucatan peninsula in Mexico as the nemesis. In *Evolutionary Catastrophes*, Vincent Courtillot promotes large-scale vulcanicity as the driving force behind this and other major extinction events.

Courtillot acknowledges the occurrence of the impact event at the 65 million-year-old Cretaceous/Tertiary (K/T for short) boundary. But he questions whether it was the impact event or, instead, massive and prolonged contemporaneous eruption of the Deccan lavas in India which really caused the extinction.

Courtillot extends the argument to the other four large extinction events of the past 450 million years of Earth's history. He points out that the K/T event is the only one of five major extinctions to have a well-established link to an extraterrestrial impact big enough to cause environmental damage on a global scale. Instead, he argues that large-scale vulcanicity is a more likely cause, not only at the K/T boundary, but also in three of the four other major extinctions, including the biggest of them all, the Permian-Triassic event. During this latter extinction, 248 million years ago, it is estimated that 95% of fossil species were wiped out over a period of two million years or so, including some 60% of marine families and 63% of tetrapod families.

Nobody now seriously denies that a major impact event occurred at the end of



New lease of life for a dinosaur prize

Como Bluff in southern Wyoming was the site of the world's first major discovery of dinosaur remains. Othniel Charles Marsh, a palaeontologist for Yale University's Peabody Museum, financed the dig and claimed a large proportion of the fossils excavated. He reunited the excavated bones and had lithographs made of them. The story of his achievements was chronicled in 1966 in a classic

book, *Marsh's Dinosaurs: The Collections from Como Bluff* by John H. Ostrom and John S. McIntosh. Most of the lithographs were also reproduced (an example, two views of a *Stegosaurus* sacrum, is shown above). The book has long been out of print, but is being reissued this month (Yale University Press, \$85), with a foreword by Peter Dodson which places the discovery in historical perspective.

the Cretaceous. And few will disagree that the ammonites, dinosaurs, pterosaurs and plesiosaurs suffered total extinction at the time, or that many other groups, such as the birds and marsupials, suffered major setbacks (around 75% extinction at the family level). However, there may be quibbles over the numbers and whether or not some dinosaurs just survived into the earliest Tertiary.

Although the impact crater is buried beneath a kilometre or so of younger sediments, geophysical measures of its size indicate that a bolide 12 kilometres wide did indeed smash through the atmosphere at around 30 kilometres per second. In the process, it excavated, pulverized and blasted into the atmosphere some 50,000 cubic kilometres of limestone, burning it to produce substantial volumes of carbon dioxide and sulphur dioxide. In a court of law, council for the defence would be hard pushed to get an acquittal for the impactor.

Comparing legal and geological debate is not as fatuous as it might at first seem. Much of the argument centres on the quality and interpretation of the forensic evidence, not to mention the persuasiveness of the advocates. Courtillot quotes Sigmund Freud's warning that "not even the most tempting probability is a protection against error".

He goes on to explain his problems with the impact as causal factor, for example the considerable difficulties involved in establishing the rates at which events happened in the long-lost past. Courtillot questions whether the seven metres of sedimentary rock that mark the K/T boundary at Mimbral in Mexico were "laid down in less than a week", as the catastrophist-impact fans claim, or over "many thousands of years", which better fits the more gradual vulcanist model.

Dinosaurs might make spectacular victims, but statistically their distribution at the K/T boundary does not stack up very well. By comparison, marine microorganisms, such as the planktonic foraminiferans, are often exceedingly abundant in marine boundary sediments. The problem is more a matter of expertise in identifying them and accurately plotting their species distribution across the boundary. In one of the best marine boundary sections, El Kef in Tunisia, the foraminiferans seem to show a stepped extinction, which starts before the iridium marker for the impact. If correct, this gradual extinction would certainly seem to support the volcanic cause. But Courtillot is admirably cautious, and calls for "further confirmation" even when evidence seems to support his corner.

One of the great scientific benefits of the K/T-boundary debate has been the focusing of effort and expertise by a wide range of researchers, from geophysicists to

micropalaeontologists, climate modellers and ballistics experts, on a very specific problem and time. Courtillot gives a well-argued taste of the debate for the general reader and has been very well served by his translator, Joe McClinton.

Like politicians under the baleful glare of a persistent TV presenter's demand for a straight 'yes' or 'no', the Earth-science community will continue to squirm on the defensive for some time over the K/T event. As Courtillot finally remarks, "no doubt the party will go on". And many secrets remain to be brought to light, so perhaps the very difficulty of geology is like that of politics, the 'art of the possible', or should it be the 'science of the perhaps'?

Douglas Palmer is at 31 Mawson Road, Cambridge CB1 2DZ, UK.

Unfinished portrait of the artist

Inner Vision: An Exploration of Art and the Brain

by Semir Zeki

Oxford University Press: 1999. 236 pp.

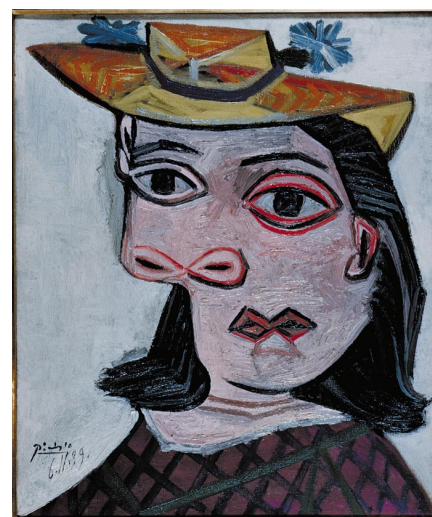
£19.99, \$35

John Nash

I discovered Semir Zeki's *A Vision of the Brain* (Blackwell Science) in 1993, browsing the science floor of my university's library. I read it with fascination and bought my own copy when it appeared in paperback. Zeki was then professor of neurobiology at University College London. His monograph is a richly detailed but lucid history of research into, and current knowledge of, the anatomy and physiology of the brain and, in particular, the visual cortex.

When I first became interested in the psychology and physiology of visual perception as an art student, soon after the Second World War, the brain was essentially a black box: its activities could only be inferred from its performance. But in the past half-century, understanding of mechanisms of the brain has been transformed. Zeki's professional concern is analysing the net of neural connections in the visual cortex. We now know that it is modular in its structure: different areas of the visual cortex respond, for example, to colour, orientation of lines, and movement.

As a painter and art historian interested in visual perception, I was intrigued to find that Zeki's new book is about painting. His thesis is that "the function of art and the function of the visual brain are one and the same"; his aim is to "develop the outlines of a theory of aesthetics that is biologically based". One of his strategies is to relate those modules of the visual cortex revealed by



Perceived truth: Picasso tried to represent the world as it is, not as it appears to the eye.

recent research to different modes and styles of pictorial representation.

I don't find that the book quite works. One reason is that, despite Zeki's astounding research into, and knowledge of, neural functioning in the visual cortex, his understanding of the psychology of human visual perception in the world, of the world, is naive, simplistic and old-fashioned. He writes: "We need to ask ... [a] question ... so obvious that it is ... never asked ... what is the visual brain for? ... The answer [is] ... we see in order to be able to acquire knowledge about this world". Never asked? Surely, it has often been noted; indeed, it was central to the theory of perception presented by the late J. J. Gibson. Zeki makes no reference to his work; here, and in *A Vision of the Brain*, his sole quotation results in a misrepresentation of Gibson's thesis.

Gibson, I am sure, would agree with Zeki (as I do) when he says that it is "the function of the brain: to represent the constant, lasting, essential and enduring features of objects, surfaces, faces, situations". Unfortunately, preceding this quotation, Zeki says "the function of art ... is very similar to the function of the brain". This is far too sweeping a statement and, in his final chapter, lands Zeki in difficulties of his own making.

Despite the illuminating sense of the history of his own science that Zeki showed in his earlier book, he shows no grasp of the importance that its history has to an understanding of painting. In his chapter, "The cubist search for essentials", he quotes Jacques Rivière, writing in 1912: "The cubists are destined ... to give back to painting its true aims, which is to reproduce ... objects as they are" (that is, not as they appear to the eye) — thus proposing that the cubists were aiming to do what Zeki claims all visual art does. But Rivière was a critic, here berating Pablo Picasso and

Georges Braque, who had been evolving forms of cubism since 1908, and others, who had been aping them since 1910, for *failing* to represent objects as they are, as he argued they should do.

In his final chapter, "Monet's brain", Zeki ties himself in knots. Claude Monet clearly does not fulfil Zeki's dictum that art should "represent the constant, lasting, essential and enduring features of objects". He is baffled by Monet's ability to represent not the constant colourings of the world but appearances under changing light conditions. Such features, Zeki claims, are visible only to brain-damaged "dyschromatopsics" (people suffering from a disorder of colour vision).

It has often been claimed, says Zeki, that Monet "painted with his eye, but, Great God, what an eye". But he wants to show that even for Monet "the higher cerebral centres played a very critical role in his work" — a truism, surely — and, also, "that his work was far from being an attempt to capture the fugitive moments". How far, exactly?

Monet's 30 canvases of the west front of Rouen cathedral under diverse light conditions, painted in 1892–94, might suggest, at a glance, says Zeki, that Monet could have been dyschromatopsic. "The suggestion is insulting if not laughable", he continues. He then identifies the areas of the visual cortex that would have been activated as Monet painted. "All this can be surmised from what happens in the brain of a normal subject when he [*sic*] views a coloured scene."

Zeki claims Monet "was, in fact, using the knowledge in his brain to deliberately paint something that departed from what he was actually seeing".

But Monet had been a professional and most prolific painter for more than 30 years when he painted the cathedral series. He had spent hours of each day studying and attempting to depict just those qualities of changing light that most of us, concerned with our own affairs, are indifferent to but which, nevertheless, are not invisible to us. Otherwise, we would not enjoy Monet's paintings. Having seen his images, we can return to the world and see it through his eyes. For Monet, his professional skill was also his curse. Looking at his wife on her death bed, he wrote: "I caught myself watching her tragic forehead, almost mechanically observing the sequence of changing colours which death was imposing on her rigid face. Blue, yellow, grey and so on."

Although Zeki tells us much about the functioning of the visual cortex, his observations on paintings are not illuminating. But do read *A Vision of the Brain*. It is truly impressive. ■

John Nash is in the Department of Art History and Theory, University of Essex, Colchester CO4 3SQ, UK.

Science's stall in the global market-place

On the Edge: Living with Global Capitalism

edited by Will Hutton and Anthony Giddens
Jonathan Cape: 2000. 241 pp. £8.99 (pbk)

Norman Myers

Every day a whopping \$1.5 trillion shuttles back and forth through the stock exchanges of the world. That's quite a slice of the \$32 trillion global economy. Moreover, there is really only one big stock exchange, as Wall Street, the City of London and their counterparts in Frankfurt, Zurich, Tokyo and so on trade around the world's clock, intensifying the mobility of electronic capital. It funds economic activities with all manner of unknown repercussions. Which of my investments underpins over-logging in Cameroon, or sweatshop factories in Bangkok, or unsustainable farming in Mexico? Whatever my reservations about the process, I am globalizing as much as the rest of us. I am closely enough involved with billions of fellow members to shake financial hands with them or tread on their toes.

The same applies to the environment. If the Chinese, sitting on one-third of the planet's coal, burned their stocks to fuel development, that could pump as much carbon dioxide into everyone's atmosphere as would be avoided were Britain and Germany to get out of fossil fuels altogether. The planetary ecosystem is a seamless continuum, and the winds carry no passports.

These are some of the thoughts prompted by *On The Edge*, published this week, with its assessment of global and hence unitary capitalism. Of the 100 biggest economic entities in the world, only half are nation states, the rest being corporations. The world is no longer governed by governments alone and our daily lives — cultural perceptions, social mores, value systems, scientific endeavours — are increasingly determined by business leaders. Where are they leading us? If we want them to change direction, how can we make the message loud enough to be heard?

These issues are explored by 13 contributors, including economist Paul Volcker, investor George Soros, physicist and philosopher Vandana Shiva and journalist Polly Toynbee. They applaud the opportunities afforded by global-scale corporations. But they also alert us to the problems of profits-driven business with its lack of social responsibility and its frequent indifference to externality costs, especially environmental costs, and other risks such as concentrated control of the media. They highlight the socio-economic apartheid of a winner-take-all system where many people will end up in exclusion.

The adverse environmental impacts have been widely documented, even though they attract little attention from international agencies. Notable examples include the inequitable use of tropical genetic resources for agriculture, and carbon dioxide emissions from the fossil fuels industry. According to the American analyst Paul Hawken, US businesses cause so many spillover costs — not just cancer from tobacco, pollution from cars and waste from manufacturing, but also the costs of community disruption — that they could cost society \$3 trillion per year, or five times their profits. Accountability, anyone?

Fortunately, a host of constraints on business practices are emerging, imposed as much by non-governmental organizations (NGOs) as by governments. In 1975 there were only 1,400 NGOs; today there are well over 30,000. These citizen activists, often globalized themselves, can deploy much muscle in the market-place: witness the 1995 campaign against Shell's plans to sink the Brent Spar oil platform. Similarly, NGOs have confronted RTZ and De Beers for supporting governments that ignore human rights.

Global capitalism has major implications for science, too. For one thing, there is a parallel globalization of the scientific community. Courtesy of the Internet and e-mail, scientists can engage in virtual conferencing with colleagues around the world. There has been a quantum advance in academic networking which transcends national boundaries and interests at the touch of a keyboard. This enables scientists to keep pace with the headlong expansion of global capitalism, and to induce scientific constraints on business activities. There will surely be many more GMO-style sagas.

For a second thing, science is increasingly pursued through the largesse of big business, notably in the form of pharmaceutical corporations with mega research budgets. A number of chemical companies control crop seeds and thus our food-chains through their biotechnologies.

Scientists are also involved through Britain's third largest pension fund, the Universities Superannuation Scheme, which, with assets of £20 billion (US\$30 billion), is to challenge businesses over their environmental performance or unduly low wages.

All in all, this is an illuminating exploration of the fast-changing landscapes of capitalism, with its multiple links to everyone's lifestyles. Primarily directed at the corporate community, it has much to say about what scientists should bear in mind as they interact, whether wittingly or not, with the wider world. The book is not only an educational read, it is entertaining as well, thanks to its vigorous debates. Read it and ponder some of the implications of pursuing science at Earth, Inc. ■

Norman Myers is at Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK.

A century of cognitive decline

If we live long enough, will we all become demented?

Bruce A. Yankner

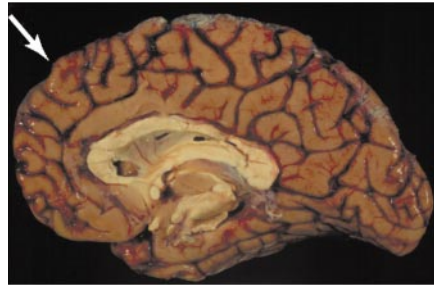
The relationship of longevity to cognitive decline was a major medical issue of the twentieth century. Several evolutionary arguments have been advanced to account for either the preservation or loss of cognitive function with ageing. For example, it has been argued that dementing illness occurs at an age that was rarely achieved during much of man's evolution, and therefore could not have been subject to selective pressure. It has also been suggested that the preservation of function in old age conferred a selective advantage by creating a class of older individuals who could care for children, freeing the mother for other activities such as foraging or food preparation.

However, an equally plausible argument is that the burden of an ageing population would have selected for rapid cognitive and physical decline after the age of maximal reproductive fitness. These arguments are all based on the idea that cognitive decline is linked in some way to the ageing process — a concept that has undergone major upheaval in the past century.

The description by Alois Alzheimer in 1907 of a distinct pathology associated with dementia laid the foundation for the modern study of neurodegenerative diseases. However, it was not until late in the century that the molecular components of amyloid plaques and neurofibrillary tangles — the hallmark lesions of Alzheimer's disease — were elucidated. An intense debate emerged as to which of these two lesions, or some other process, is the cause of dementia.

Proponents of the so-called amyloid hypothesis argue that all genetic mutations that give rise to Alzheimer's disease also predispose to amyloid deposition, and that amyloid is neurotoxic. Sceptics note that amyloid plaques correlate poorly with the degree of dementia, and that animal models of amyloid-plaque deposition do not replicate the neurodegenerative changes in an Alzheimer brain. The debate rages on, and will probably not be settled until drugs are available that block plaque formation.

The focus on plaques and tangles has obscured a potentially more fundamental issue — the relationship of dementia to the ageing process. Although there are rare examples of early-onset dementia due to genetic mutations, the vast majority of dementing illness occurs after the age of 65. For much of the century, Alzheimer's disease was thought of as senility, an inevi-



Our brains shrink as we age: the brains of normal 60- and 98-year-olds (top and middle, respectively), and a 76-year-old with Alzheimer's disease (bottom; not to scale). There is age-related atrophy of the frontal lobe in the 98-year-old (arrows), and extreme atrophy in Alzheimer's disease. Photo courtesy of Neil Kowall.

table decline in the cognitive ability of aged individuals.

It took the ground-breaking studies of Robert Terry, Henry Wisniewski and their colleagues in the 1970s to show that dementia of the Alzheimer type is a disease, rather than an inevitable consequence of ageing. Plaques and tangles can appear in the brains of many cognitively normal aged individuals, but the degree of pathology is usually much greater in Alzheimer's disease. This watershed concept argued against the idea that senility is normal, and gave rise to the hope that Alzheimer's disease could be prevented or reversed.

Although Alzheimer's disease is not a consequence of normal ageing, it is likely to be related in some way to the ageing process.

Most people over the age of 70 exhibit some degree of cognitive decline, particularly for short-term memory, and this accelerates with age. Furthermore, a new clinical entity, called mild cognitive impairment, has been increasingly recognized in the elderly.

This syndrome usually consists of an isolated memory deficit which is not severe enough to fulfil the criteria for Alzheimer's disease, but is greater than average for age. Some, but not all individuals with mild cognitive impairment will progress to Alzheimer's disease. Even asymptomatic individuals who carry the apolipoprotein E4 allele, a genetic risk factor for Alzheimer's disease, can show reduced brain activity in PET scans many years before the onset of cognitive decline. Thus, Alzheimer's disease may be preceded by a long, clinically silent prodrome.

The continuum between normal ageing, mild cognitive impairment and Alzheimer's disease provides support for an old idea, namely, that Alzheimer's disease may be an accelerated form of brain ageing, and raises an old conundrum: if we live long enough, will we all become demented?

Aside from the philosophical issues, this concept has important implications for understanding the mechanisms of neurodegenerative diseases. If these diseases represent pathological variants of the ageing process, then understanding mechanisms of normal brain ageing may lead to fundamental insights into the disease mechanisms.

For example, Alzheimer's disease, Parkinson's disease and frontotemporal dementia are all characterized by abnormal accumulations of aggregated proteins. Are these disorders extreme manifestations of the impaired protein folding that occurs to a lesser extent in normal brain ageing? If so, would therapeutic interventions for these disorders also retard normal age-related cognitive decline? Conversely, would therapeutic interventions designed to ameliorate age-related protein misfolding in the brain be effective in a variety of neurodegenerative disorders? The twentieth century has witnessed a dramatic prolongation of lifespan, but little progress in preventing age-related cognitive decline. The anticipated further prolongation of human lifespan in the twenty-first century will be a hollow victory unless cognitive function can also be preserved. ■

Bruce A. Yankner is in the Department of Neurology, Harvard Medical School and Division of Neuroscience, Children's Hospital, Boston, Massachusetts 02115, USA.

The song of the Neanderthal

From ancient DNA, to an unusual nose, to a sound you can't quite place.

Mark W. Tiedemann

André Gilles-Rasson is emblematic of the staff at the PaleoGenome Retrieval Project. Casual, sincere, quietly optimistic, he gives the impression of a personality consultant rather than a researcher on the cutting edge of palaeolithic bio-mapping.

"I suppose we'll never be rid of that label," he says, smiling ruefully. "My predecessor used it in one interview and we've been stuck with it ever since."

"It's inaccurate?"

"Distracting. You know what our work is here?" He gestures out of the broad window, beyond which a series of bioisolate environmental domes stretches towards the distant Pyrenees.

"Rainforest recovery and preservation."

"Among other things, yes. But we aren't recreating Stone Age hominids."

I hold up the disc that has just been released by the project's Community Interface Office, with its cover depiction of a stylized Neanderthal sporting a startlingly complex nose. He grins and nods.

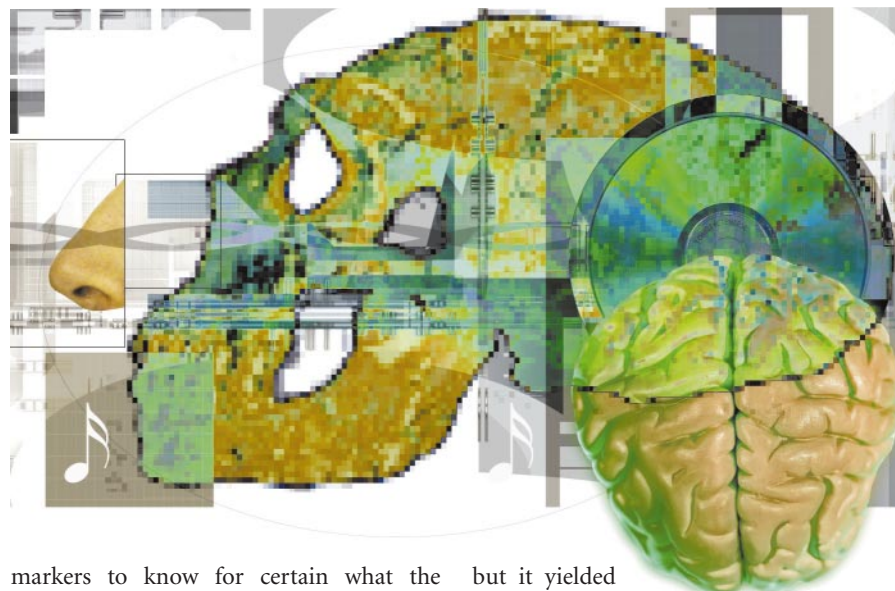
"You want to know how we recorded Neanderthal music? You are aware that we originally did pure pharmaceutical research here? We still do, but there is so much more now. Many of our products are psychotropics — antidepressants, bipolar remedies, treatments for epilepsy and so forth. Consequently, we developed methods for mapping the brain that gave us accurate, reliable modelling of how the brain is affected by different molecules.

"We joined up early in the century with an offshoot of the Human Genome Project and began working on new systems for projective and holistic modelling. First with humans, then with other things. Take any sample of DNA and use it to model the organism it is programmed to express. The more complete the sample, the better the results, of course, but we have become quite adept at the extrapolative side, filling in the blanks, so to speak.

"About 22 years ago we began working with a team of palaeobiologists on fossil DNA."

"Dinosaurs?"

"Some. We've been able to validate certain theories about them. Most of our successes have been with plants. Time is the factor — the nearer to the present, the more complete the DNA. In some instances we have enough trace remnants of the protein



markers to know for certain what the sequences are."

"As with Neanderthals?"

"We were very, very lucky. We've been able to model an entire organism, including — and this was the real stroke of luck — the brain."

"Forgive me, but that nose stands out."

"That was a surprise. But it would still be just a nose — an unusual one — without the synaptic structure of the brain to tell us how the Neanderthals used it."

"The music."

Gilles-Rasson nods, leaning forward. "Soft tissue — muscle, skin, cartilage — does not fossilize well. Even in those rare instances where tissue has survived, we're talking about specimens ten- to twenty-thousand years old. When the gene map produced this structure, we were totally surprised. The trick, then, was to figure out its function."

"According to the liner notes, Jean-René Compiègne suggested communications?"

"He suggested music right off. He and Marie Conéal actually built a physical model of the nose to test it. It's amazingly versatile."

"And the music itself?"

He sighs. "Music. A puzzling psychological phenomenon. All the theories added together fail to explain it in Darwinian terms. Infants with no prior experience of it respond, and the response is different from the response to simple communication. There is an emotional reaction at odds with simple informational modes. Once we had the synaptic structure and the instrument, we were able to make maps of the emotional range — anticipate how Neanderthals would react to music.

"Well, anticipate ... frankly, we guessed ...

but it yielded some fascinating results. When we understood the patterns our models showed us and matched those to the tonal and textural capacities of that nose, we realized that we might have a basis for deriving compositions."

"It's an amazing recording. I don't know how to classify it."

"It's a kind of whole-tone scale, but the harmonies are based on fourths instead of thirds, so there is an eeriness to it. Not haunting, but familiar. *Presque vu, non?* Like something we should have known, something we might once have heard, had we listened."

"I wonder what Stravinsky would have said to it."

"Perhaps 'But of course!' Or perhaps nothing. We've found that a significant number of people simply do not hear it as music."

"I've noticed that among my own friends and colleagues. They don't dislike it, as they might music not to their taste. They simply don't 'hear' it. Can you explain that?"

He shrugs. "Would we recognize the music of someone from Tau Ceti as music? Different species. The synaptic structure, the way the Neanderthal brain worked — our models show that it was quite different from ours in many respects. They must have possessed a completely different aesthetic."

He smiles slyly. "Of course, we may not really know what the true music of *Homo sapiens*. That may be our next discovery."

(Interview from *Aurikle News and Review* March 2089, reproduced by permission.) ■

Mark W. Tiedemann's first novel, *Mirage*, is a spring release from iBooks. Volume One of his *Secantis Sequence* will be published by MeishaMerlin in the autumn of 2001. He lives in St Louis, Missouri.

Life's downs and ups

Douglas Erwin

The peak in recovery of biodiversity seems to lag the peak of an extinction by about ten million years. This pattern is independent of the severity of extinction, implying that recoveries create new ecological opportunities.

Ecology is in part an experimental science: by removing species from or otherwise perturbing present-day ecosystems, ecologists can test competing models of the way in which organisms are affected by each other and by their environment. In investigating past episodes of extinction, and the biotic recovery from them, palaeontologists lack that luxury. They are also dealing with a far larger setting, the fossil record of the Phanerozoic, which runs from 543 million years ago to the present. However, the observable Phanerozoic fossil record does, in effect, provide replicate experiments. At least five mass extinctions and subsequent recoveries in biodiversity occurred during this time, as well as many smaller biotic crises and recoveries, allowing a comparison of similar (if not identical) events in a search for general patterns.

On page 177 of this issue, Kirchner and Weil¹ analyse this record. They show that for extinctions at all scales (that is, whatever the severity of the extinction), the peak origination rates of new taxonomic groups lag extinction pulses by some ten million years. This result stands in sharp contrast to expectations that an extinction event will be followed by an immediate pulse of originations,

and to empirical data suggesting that most recoveries from mass extinctions are complete in less than five million years².

Palaeontological approaches to biotic recoveries have been heavily influenced by equilibria-based models derived from island biogeography. According to this view, extinctions perturb the equilibrium, and recoveries represent a refilling of available ecosystem niches and a return to equilibrium^{3–5}. Under many models, the refilling should follow a logistic curve (Fig. 1) and be exponential, so the greater the extinction, the longer the recovery period. (These equilibria can break down under sufficient pressure, however. The pervasive extinction at the end of the Permian, 251 million years ago, seems to be one case in which the perturbation was so great that the system was evidently reset⁶.)

There are far fewer studies of post-extinction recoveries than of mass extinctions. Those that exist reveal considerable variation in patterns of recovery, but several generalities emerge² (see Box 1). In the immediate aftermath of mass extinctions, biodiversity is very low, with the dominance of geographically widespread generalist organisms adapted to a wide variety of eco-

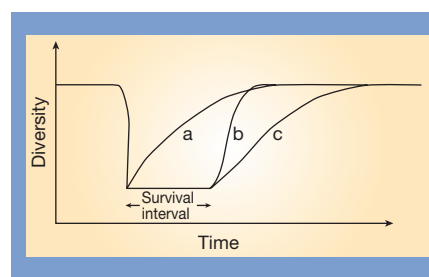


Figure 1 Different possibilities for the recovery in biodiversity after extinctions. Curve a shows immediate logistic growth, as expected from many models of the process. Curve b consists of a lag, followed by rapid origination of taxonomic groups in which, by positive feedback, new species create new roles for yet more species. In option c there is logistic growth after a lag during the 'survival interval', the time during which biodiversity remains low, with a predominance of generalist organisms that are adapted to a wide variety of ecological conditions. The analysis of Kirchner and Weil¹ favours option b.

logical conditions. This is the 'survival interval'. It is followed by the rapid appearance of new taxonomic groups during a phase of swift rediversification, in which limited data suggest that there might be considerable regional differences⁷. Survival intervals have highly variable durations, and no consistent pattern emerges between the magnitude of extinction and the pattern of recovery, or the extent of change in ecological communities.

Kirchner and Weil¹ have taken a different approach to the problem. They do not concentrate on major extinctions. Rather they have taken the late Jack Sepkoski's compendia of the fossil-record ranges of two marine taxonomic groups — families and genera — during the Phanerozoic. They have then cross-correlated peaks in extinction and origination throughout that record to determine the similarity between the two time series.

This is no easy task. The division of the stratigraphic record into stages of variable, and uncertain, duration, and the dependence of diversity at one stratigraphic interval upon diversity in the preceding interval (in statistical terms, auto-correlation), raise challenging problems. Moreover, because many of the smaller 'extinctions' would in fact be expected to be random fluctuations in the quality of the fossil record, one might

Box 1 Defining recoveries in biodiversity

Extinctions are easily defined: the disappearance of a group, or groups, of organisms.

Complications can arise in particular cases owing to the variable quality of the fossil record, but the definition is straightforward and the problems in analysing extinctions are tractable.

Post-extinction recoveries in biodiversity are far more difficult to tackle². The most common definition takes in a brief, low-biodiversity 'survival interval' immediately after the extinction, followed by a phase of rapid rediversification, and finally a decline in origination rates of taxonomic groups to normal values. Complications arise when recoveries differ in

duration between groups, or between regions, after the same episode of extinction.

Palaeoecologists can also focus on recovery as the period leading to the return of normally functioning ecosystems, but it is not easy to apply this principle in practice. Perturbations of the carbon cycle are reflected in changes in carbon isotopes; taking such geochemical data as proxies for ecosystem function, a recovery is complete when isotopic values return to pre-extinction values or a new equilibrium is established⁹.

Kirchner and Weil¹ adopt a very different definition: they define the 'recovery interval' as

the time between a peak in extinction rate and the subsequent peak in origination rates, implicitly assuming a link between the two peaks.

In general, the recovery of geochemical proxies is the quickest (under one million years), followed by recovery to the palaeoecological and diversification-based definitions (between one and five million years; ref. 2), then recovery to Kirchner and Weil's definition (about ten million years). As more-detailed studies of post-extinction rebounds in biodiversity are conducted, the relationships between these different perspectives should become apparent.

D.E.



100 YEARS AGO

A curious example of resonance is to be noticed in Llandinat Church, Llandovery, South Wales. In one of the windows there is a pane of glass which is not very tightly fixed, being free to oscillate with a definite frequency, which happens to correspond to the frequency of the low pedal "G" of the organ. The consequence is that when the service is taken in G, at the end of each of the Responses, Amens, &c., quite a loud buzzing noise is produced by the resonance of the window; and I have seen strangers sitting near the window seem quite perplexed, not knowing what causes the noise.

From *Nature* 8 March 1900.

50 YEARS AGO

Among the diseases which fill the out-patient departments of hospitals and dispensaries in tropical Africa, yaws and tropical ulcer are undoubtedly the most time-consuming. Up to the present, yaws has been treated solely by injections of bismuth, arsenicals or penicillin.... The following observations show that aureomycin given by mouth is effective in clearing up the secondary lesions in yaws. Three children all with florid secondary yaws were given aureomycin, six capsules (250 mgm. per capsule) daily for seven days. As African parents consider that yaws treatment without injections cannot possibly be effective, the three children were also given a daily injection of 2 ml. of physiological saline as a placebo. No change in the lesions was noted until the fifth day of treatment, and treponemata could still be seen in the lesions up to this time. On the fifth day the lesions began to dry up and by the seventh day they had almost completely healed. The children gave positive Kahn reactions at the time of beginning treatment and continued to show positive Kahns for six weeks after ceasing treatment, although they remained in good health and exhibited no relapse. Thus, as in syphilis, so in yaws, aureomycin brings about rapid healing of the lesions, but in the doses so far used no serological changes have been noted.... Aureomycin is somewhat slower in action in penicillin; but it has the advantage of not requiring intramuscular injections. It would thus be simpler to carry out mass treatments with aureomycin than with penicillin.

From *Nature* 11 March 1950.

expect that the results would be gibberish, or would be overwhelmingly influenced by the well-known mass extinctions and smaller biotic crises.

Kirchner and Weil carefully account for these difficulties. In contrast to the evidence from recoveries in biodiversity after mass extinctions, their analysis shows no immediate burst of origination of new taxonomic groups after extinctions. On average, peak origination rate lags peak extinction rate by ten million years. Although much of the support for this conclusion comes from the five mass extinctions, the cross-correlation remains statistically significant even after removing them from the data.

This is not the prediction of a logistic model, and if these results¹ are confirmed they imply that recoveries in biodiversity follow a very different process from the filling-in of empty ecological niches. Recoveries seem to have their own dynamic, and the authors plausibly suggest that it reflects a positive feedback process, with new species creating roles for yet more species (Fig. 1). The loss of only a few species might leave niches that can be quickly refilled by closely related taxonomic groups, but larger extinctions seemingly cause a disappearance of the ecological fabric, which can only slowly be rebuilt. Just such a pattern of recovery has been proposed for equatorial forests after the end-Permian mass extinction⁸.

The pattern documented by Kirchner and Weil could still be consistent with a logistic model if the survival interval scales to the magnitude of the extinction. Larger extinctions would delay the onset of the exponential increase, producing a scale-independent lag. Because the reliability of the radiometric dates on the geological timescale is still so low, and the durations of recoveries are so poorly known, this possibility is also consistent with Kirchner and Weil's results. We are clearly a long way from understanding the dynamics of post-extinction rebounds in biodiversity. But detailed, comparative studies at a variety of scales promise to provide insights into recoveries from a range of biotic disturbances — and, perhaps, a glimmer of our own impact on the course of evolution. ■

Douglas Erwin is in the Department of Paleobiology, MRC-121, National Museum of Natural History, Washington DC 20560, USA.
e-mail: erwin.doug@nsmnh.si.edu

1. Kirchner, J. W. & Weil, A. *Nature* **404**, 177–180 (2000).
2. Erwin, D. H. *Trends Ecol. Evol.* **13**, 344–349 (1998).
3. Sepkoski, J. J. Jr *Paleobiology* **10**, 246–267 (1984).
4. Benton, M. J. *Trends Ecol. Evol.* **12**, 490–494 (1997).
5. Alroy, J. in *Biodiversity Dynamics* (eds McKinney, M. L. & Drake, J. A.) 232–287 (Columbia Univ. Press, New York, 1998).
6. Van Valen, L. M. *Nature* **307**, 50–52 (1984).
7. Jablonski, D. *Science* **279**, 1327–1330 (1998).
8. Looy, C. V., Brugman, W. A., Dilcher, D. L. & Visscher, H. *Proc. Natl Acad. Sci. USA* **96**, 13857–13862 (1999).
9. D'Hondt, S., Donaghy, P., Zachos, J. C., Luttienberg, D. & Lindinger, M. *Science* **282**, 276–279 (1998).

Quantum cloning

Schrödinger's sheep

Wojciech H. Zurek

Dolly, the sheep cloned by using DNA extracted from her older 'identical twin', caused a scientific sensation a few years ago. Information stored in a DNA molecule can be copied, and, more generally, manipulated in a manner essentially analogous to the records contained in computer files or in any other memory device. Making or deleting a copy of an existing file is a privilege we take for granted. Yet, on page 164 of this issue, Pati and Braunstein¹ argue that deletion of unknown quantum states is not possible, a result which complements the 'no-cloning theorem' — a prohibition on copying of unknown quantum states^{2,3}.

Restrictions on cloning and deleting quantum states add a new twist to an old question: if quantum theory is universally valid, how can it be consistent with the experimental reality (here, of sheep cloning)? Moreover, if an unknown quantum state cannot be copied or manipulated, what does it mean for a quantum system to be in a certain state?

First, a distinction between outright era-

sure and the more subtle operation of deletion should be noted. Erasure of either classical or quantum information — getting rid of the very last copy of the data — is thermodynamically irreversible. This was anticipated by Leo Szilard⁴ but fully understood only by Rolf Landauer⁵, who realized that logical irreversibility — the impossibility of deducing the past from the present following erasure — begets thermodynamic irreversibility. Replacing a jumble of data bits (0's or 1's) with a perfectly ordered 'blank' state (that is, a sequence of only 0's) requires a reduction of disorder, and is therefore thermodynamically costly.

Deletion in the narrow sense used by Pati and Braunstein¹ is a different beast. Deleting an extra copy of a classical bit is logically reversible and has no thermodynamic cost — one can use the original still present in the memory to reconstruct a new copy. Such deletion could start with two arbitrary and unknown, but identical bits — an original and a copy. Controlled not (C-NOT for brevity) is the conditional logical operation

Box 1 No cloning or deleting

Why can't you delete an unknown quantum state? Consider a quantum C-NOT which does the right thing for every pair of $|original\rangle$ and $|copy\rangle$ logical states, taking $|0\rangle|0\rangle \rightarrow |0\rangle|0\rangle$, $|0\rangle|1\rangle \rightarrow |0\rangle|1\rangle$, $|1\rangle|0\rangle \rightarrow |1\rangle|1\rangle$, $|1\rangle|1\rangle \rightarrow |1\rangle|0\rangle$. When presented with a pair of

qubits which are in a general state $|s\rangle|0\rangle$, quantum C-NOT will create a copy of the original only when $|s\rangle$ is either $|0\rangle$ or $|1\rangle$. In the case of a general $|s\rangle$, this C-NOT will entangle the original and the copy, producing $(\alpha|0\rangle + \beta|1\rangle)|0\rangle \rightarrow (\alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle)$.

This is not cloning. As the reader can verify by starting with a pair of general but identical states $|s\rangle|s\rangle$, deletion won't work either. In particular, when the original and the copy start in the complementary states $|\pm\rangle = (|0\rangle \pm |1\rangle)$ the original will get deleted! **W.H.Z.**

used to reset the copy bit to 0. C-NOT 'flips' the state of the copy when the original is 1, and does nothing otherwise. So when C-NOT is presented with two identical 1's it will leave the original bit untouched, but reset the copy to 0, resulting in deletion. Moreover, C-NOT is logically reversible — repeating it ensures that the 0 on the blank copy becomes 1 when the original bit is 1, but stays 0 otherwise. Last, but not least, C-NOT will do its job of deleting whether or not the original and copy are 'known'.

Why can't one repeat the same trick in the quantum domain? Quantum C-NOTs exist and are reversible. So why can't one delete an extra copy of an unknown quantum state? The problem can be traced back to the difference between quantum and classical ignorance. Indeed, it touches on the very nature of the quantum state. In quantum physics, when dealing with a quantum two-state system (a qubit) we can be ignorant of it in a way that is quite distinct from what happens classically. A qubit can be in an arbitrary superposition of $|0\rangle$ and $|1\rangle$, the two states that correspond to the classical 0 and 1. So the state $|s\rangle = \alpha|0\rangle + \beta|1\rangle$ (where α and β are complex numbers and $|\alpha|^2 + |\beta|^2 = 1$) is a perfectly legal quantum state. But the actions of the quantum C-NOT imitate the classical C-NOT only when the copy and the original are in the preferred states, $|0\rangle$ or $|1\rangle$.

As Pati and Braunstein¹ demonstrate, the problem with deleting is similar to that with cloning. An operation that can be used to clone or delete a state (for example C-NOT) is guaranteed to work only for a preferred set of mutually exclusive (or 'orthogonal') quantum states, such as $|0\rangle$ and $|1\rangle$. But when applied to superposition states of the preferred quantum states the same C-NOT will fail to clone or delete (Box 1). The problem with quantum deletion or cloning is then ultimately traced back to the principle of superposition. An unknown classical state can be safely assumed to be one of a limited set of distinct alternatives: 0 or 1 for the bit. An unknown quantum state can be one of many arbitrary superpositions (such as $|s\rangle$ for a qubit) of these alternatives.

One could still clone or delete, provided one knew $|s\rangle$. First, one would need a rota-

tion to align $|s\rangle$ with $|1\rangle$, follow it with the usual C-NOT, and then undo the rotation. But when $|s\rangle$ is not known, the appropriate rotation cannot be selected. Moreover — and this is the other big problem with quantum states — in contrast to classical information, an unknown quantum state cannot be measured without being disturbed (for more or less the same reason that it cannot be cloned). In effect, one objective of measurement is to produce a copy — to 'clone' the information contained in a quantum state. This is a classic Catch-22. To clone or to delete, one needs to know the state, and one cannot know it unless one measures (that is, 'clones') that state.

But Dolly did get cloned. So have we found an experimental falsification of the universal validity of quantum theory? Will Dolly join Schrödinger's cat (which should exist in a superposition of dead and alive states) as a famous quantum paradox? Of course Dolly is classical. And so are the cats, but this does not resolve the question: DNA is, after all, made out of flagrantly quantum atoms. So why did the quantum prohibition against manipulating information get suspended for sheep DNA?

At least we have an answer to this last quandary. Information in DNA or comput-

ers can be accessed and manipulated without being disturbed because, even in the microscopic domain, the range of quantum states becomes restricted to a few mutually exclusive possibilities. This is known as environmentally induced superselection, or einselection. It arises as a consequence of the interaction between the system and the environment, so-called decoherence, and can be explained in the language of C-NOTs (ref. 6). So a strand of DNA, at any given location, has a definite small molecule corresponding in effect to either 0 or 1, and never a superposition of two different molecules. Genetic information can be safely copied because it is never encoded in superposition states — they would be instantly eradicated by decoherence. This process forms a necessary link between the quantum and classical worlds.

It is impossible to clone perfectly. But as research in recent years has demonstrated, one can create approximate clones all of the time, or perfect clones some of the time^{7–12}. It is natural to ask whether the same might be true for deletion. Investigating strategies for approximate deletion might shed further light on its relation with quantum cloning.

Wojciech H. Zurek is in the Theory Division, T-6 MS B288, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.
e-mail: whz@lanl.gov

1. Pati, A. K. & Braunstein, S. L. *Nature* **404**, 164–165 (2000).
2. Wootters, W. K. & Zurek, W. H. *Nature* **299**, 802–803 (1982).
3. Dieks, D. *Phys. Lett. A* **92**, 271–272 (1982).
4. Szilard, L. *Z. Phys.* **53**, 840–856 (1929).
5. Landauer, R. *IBM J. Res. Dev.* **3**, 183–191 (1961).
6. Zurek, W. H. *Phil. Trans. R. Soc. Lond. A* **356**, 1793–1821 (1998); <http://xxx.lanl.gov/abs/quant-ph/9805065>
7. Bužek, V. & Hillery, M. *Phys. Rev. A* **54**, 1844–1852 (1996).
8. Duan, L.-M. & Guo, G.-C. *Phys. Rev. Lett.* **80**, 4999–5002 (1998).
9. Massar, S. & Popescu, S. *Phys. Rev. Lett.* **74**, 1259–1262 (1995).
10. Gisin, N. & Massar, S. *Phys. Rev. Lett.* **79**, 2153–2156 (1997).
11. Bruss, D., Ekert, A. & Macchavello, C. *Phys. Rev. Lett.* **81**, 2598–2601 (1998).
12. Cerf, N. J. *Mod. Opt.* **47**, 187–192 (2000).

Neurobiology

A chorus line

Emilio Salinas and Ranulfo Romo

Our senses are continuously being bombarded by a plethora of inputs. When one of them is deemed important, attention allows us to concentrate on it and shields us from the irrelevant ones. What happens in the brain when we pay attention to one object and ignore others? This is the question tackled on page 187 of this issue by Steinmetz *et al.*¹, who suggest that attention makes neurons join forces to reinforce their representation of the relevant object.

Each neuron in the brain normally receives messages sent to it by thousands of other neurons. These messages are transmit-

ted in the form of discrete electrical pulses, or 'spikes'. It is generally accepted that the number of spikes fired by a neuron per unit time conveys useful information to other neurons. This strategy, known as a 'rate code', is consistent with the simple notion that neurons act as forgetful spike counters: a neuron fires if enough input spikes arrive within a certain time window, regardless of their fine temporal pattern. However, in principle, additional information can be encoded in other ways that are more subtle — and potentially much more powerful — precisely by conferring some meaning to the times

that elapse between spike arrivals. Steinmetz *et al.*¹ provide evidence that spike synchrony, one specific form of temporal pattern, correlates with changes in attentional state in monkeys carrying out specific tasks.

This work springs from the following question. What would happen if pairs of neurons that normally send independent messages started to synchronize them, forming a chorus of spikes? There could be two (or more) consequences. For instance, synchrony might increase the loudness of the message, making it easier for other neurons to listen to it above a noisy background. On the other hand, the fact that there is a chorus might in itself be interpreted as an additional message.

The potential functions of synchrony, and neural codes based on spike timing in general, have been intensely debated^{2,3}, but experimental data on the subject^{4–6} have flowed slowly. One problem is that codes based on the temporal structure of emitted spikes may take a variety of forms, so it is not clear exactly what to look for. Another problem is that, to study codes based on synchronized spikes from several neurons, one needs to record from many individual neurons at the same time, which is technically difficult. Furthermore, the analysis of such records requires great care.

Steinmetz *et al.*¹ skilfully overcome these obstacles, and demonstrate a link between synchrony and attention. They recorded the activity of neurons in secondary somatosensory cortex (SII), an area of the brain with cells that react to touch, but the responses of which also depend on the focus of attention. These neurons typically react more to a given tactile stimulus when the animal is attending to it than when it is being ignored⁷. In all these experiments, visual and tactile stimuli were presented simultaneously to monkeys, which could perform either a visual or a somatosensory task, depending on a cue at the beginning of each block of trials. Nothing else changed across conditions, including the animals' motor responses. Crucially, the difficulty of the tasks was adjusted so that, to discriminate accurately between the tactile stimuli, the monkeys had to pay attention to them and ignore the visual display, and vice versa during the visual task. Otherwise, mistakes occurred, indicating that attention was misdirected.

Having recorded from hundreds of pairs of neurons while monkeys performed these tasks, Steinmetz *et al.* analysed the spike trains and observed three effects. First, 80% of the neurons consistently fired different numbers of spikes in the two tasks, so their firing rate was modulated by attention. This agrees with previously reported effects of attention⁷. Second, 66% of all neuron pairs were significantly synchronized during visual or tactile discrimination. This is not surprising either, because neurons that are

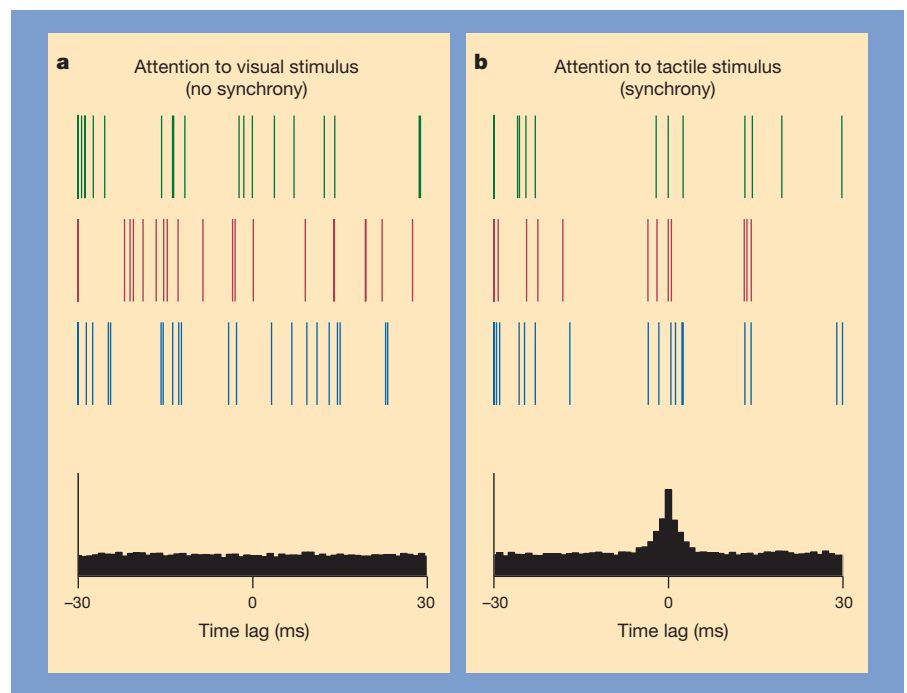


Figure 1 Neuronal synchrony in the somatosensory cortex depends on the focus of attention. Spike trains from three hypothetical neurons are shown. According to Steinmetz *et al.*¹, the degree of synchrony changes as attention is switched from one stimulus to another. **a**, When spike trains are asynchronous, the times at which one neuron fires are independent of the times at which other neurons fire, and the cross-correlogram (bottom graph) is flat. This function plots the probability that one neuron fires x milliseconds before or after another neuron has fired; x is called the time lag. **b**, When spike trains are synchronous, the cross-correlogram shows a peak, and different neurons tend to fire at around the same times.

interconnected or connected to the same inputs should exhibit some degree of synchronization. Indeed, Steinmetz *et al.* interpret this finding as a consequence of overlap between the patches on the skin surface from which individual SII neurons can be activated.

The third and most important finding is that the degree of synchrony between pairs of neurons was higher during tactile discrimination than during visual discrimination: about 99% of all SII neuron pairs fired more synchronously when monkeys paid attention to the tactile stimuli (Fig. 1). And more difficult somatosensory tasks produced larger changes in synchrony with respect to the same visual task. These results cannot be explained by a fixed pattern of neuronal connections. Nor can they be explained in terms of variations in the spike firing rates of SII neurons, because the analysis corrected for this and because sometimes synchrony increased even when the rates did not change. Whatever the underlying mechanism, these results are strong evidence that neuronal spike synchrony correlates with a cognitive process, namely, shifting of the attentional focus.

It is tempting to think that an increase in synchrony in SII can improve the processing of tactile information at the next stage in the brain. This takes us back to the question of what a chorus of spikes is good for. There is

no definite answer yet, and even theorists disagree^{2,3}. The acid test for theories that confer a function on synchronous spikes is to manipulate (for example, eliminate) synchrony while leaving everything else the same, including the firing rates. As attention affects firing rate as well as synchrony, this may never be possible. But theoretical work, computer simulations and more experiments such as those of Steinmetz *et al.* should help us figure out the extent to which synchrony influences normal cognitive function.

Emilio Salinas is in the Computational Neurobiology Laboratory, Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

e-mail: emilio@salk.edu

Ranulfo Romo is in the Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico.

e-mail: rromo@ifisiol.unam.mx

- Steinmetz, P. N. *et al.* *Nature* **404**, 187–190 (2000).
- Singer, W. & Gray, C. M. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
- Shadlen, M. N. & Newsome, W. T. *Curr. Opin. Neurobiol.* **4**, 569–579 (1994).
- DeCharms, R. C. & Merzenich, M. M. *Nature* **381**, 610–613 (1995).
- Riehle, A., Grün, S., Diesmann, M. & Aertsen, A. *Science* **278**, 1950–1953 (1997).
- Dan, Y., Alonso, J. M., Usrey, W. M. & Reid, R. C. *Nature Neurosci.* **1**, 501–507 (1998).
- Hsiao, S. S., Johnson, K. O. & O'Shaughnessy, D. M. *J. Neurophysiol.* **70**, 444–447 (1993).

Condensed matter

Too hot to melt

A. Lindsay Greer

Melting is a familiar process not expected to show surprises — the melting of ice in a cocktail is expected to produce cooling not heating. Yet just such an effect — inverse melting — has been seen in a polymeric system, as reported by Rastogi *et al.*¹ in *Macromolecules*. Their findings are relevant for a wide range of systems (not just polymeric ones), and have implications not only for the structural changes involved in melting, but also for the formation of solid amorphous (or glassy) states.

In interpreting their results, Rastogi *et al.* have revisited much earlier work^{1,2}. In his classic *Kristallisieren und Schmelzen* (1903), the great inorganic chemist Gustav Tammann devoted some 20 pages to melting, including speculation that it might sometimes be inverted³. In later works^{4–6} he reduced the emphasis on inverse melting — presumably through lack of experimental support — but retained the key concept. Liquids typically have higher volume per unit mass, but also higher compressibility, than their crystalline counterparts; so, at higher pressure, a liquid may have the same volume per unit mass as a crystal. Similarly, liquids typically have higher heat content (enthalpy, H), but also have higher specific heat, than the corresponding crystals; so, at lower temperature, the enthalpy difference may become zero.

Tammann plotted what he called a universal phase diagram for crystalline and liquid/amorphous phases (Fig. 1). He started by drawing the lines along which the volume and enthalpy differences vanish ($\Delta V = 0$, $\Delta H = 0$), as well as the phase boundary between the crystal and liquid or amorphous phases — that is, the locus of the melting temperature, T_m , as a function of pressure, P . In Fig. 1, normal melting occurs in region I, whereas in region II the sign of dT_m/dP is reversed (as observed when ice melts). Region III shows inverse melting. The crystal–melt transition, unlike the liquid–vapour transition, can have no critical point beyond which the two phases become indistinguishable. So the final, looped shape of the melting curve was attractive to Tammann because it avoided extending the curve to infinite temperatures. Later commentators⁷, and perhaps Tammann himself, considered that inverse melting would be observable only under extreme conditions. Yet, as Rastogi *et al.*^{1,2} show, such melting can be seen at moderate temperatures and pressures.

Rastogi *et al.* have studied phase transitions in the polymer poly(4-methylpentene-

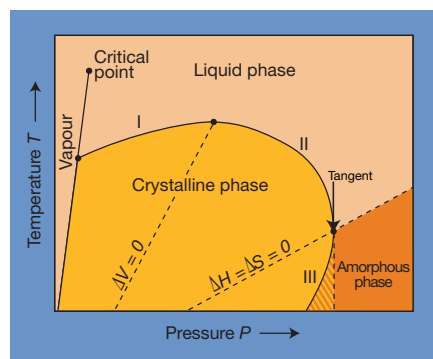


Figure 1 Tammann's universal pressure and temperature phase diagram^{5,6}. The main feature, the melting curve (T_m versus P), separates the crystalline phase from the 'melt' region, which is further subdivided into liquid and amorphous phases. Tammann noted that $dT_m/dP = 0$ when $\Delta V = 0$, and that $dT_m/dP = \infty$ when $\Delta H = \Delta S = 0$ (where V is specific volume, H is enthalpy and S is entropy). Melting as usually observed occurs in regions I and II, whereas 'inverse melting' as observed by Rastogi *et al.*^{1,2} occurs in region III. Also shown is the critical point at which the liquid and vapour states become indistinguishable.

1), P4MP1. The latest work¹ provides calorimetric confirmation for unusual phase transitions, in agreement with their previous structural studies using X-ray diffraction². The temperature–pressure phase diagram of P4MP1 (Fig. 2) shows that the normal 'tetragonal' crystal phase of this polymer is replaced by a 'hexagonal' crystal phase at high pressure. But the main interest here is in the amorphous phase, which they show to be structurally continuous with the liquid¹. Transitions to and from the amorphous phase have been studied by varying the pressure and temperature (Fig. 2).

Many systems, from metals to silicon, can be converted from crystalline to amorphous within the solid state. Rastogi and colleagues' results are notable because the transition is reversible, and occurs under near-equilibrium conditions. Because the amorphous phase is expected to be thermodynamically continuous with the liquid, the transition represented by the dashed line in Fig. 2 is inverted: an equilibrium crystallization occurs on heating, not on cooling. The pressure-induced transition (arrow A in Fig. 2) to the amorphous state (akin to melting) is found to release heat, not absorb it as in normal melting. Similarly, crystallization (arrow B in Fig. 2) is found to absorb heat, not release it as is normal.

Inverse melting, if found more generally,

might require a radical redesign of cocktails, but there is no problem here for the underlying thermodynamics. Even with inversion it remains true that the transition that occurs on heating absorbs heat (as does normal melting) and that the phase in equilibrium at higher temperature has the higher disorder, or entropy, S . What is startling is that, as a result, the crystalline phase appears to be more disordered than the amorphous phase. Analogous results, although rare, have been found in widely differing systems and in each case the greater order of average atomic positions in the crystal is offset by greater disorder in some other characteristic. For example, in colloidal⁸ and liquid-crystal systems the crystal can have greater rotational disorder of the molecules. In metallic alloys, inverse melting is possible because the crystal can have much less chemical short-range order than the amorphous phase⁹. In the case of P4MP1, the more open structure of the crystal must contribute to its overall higher entropy.

The line where entropy differences vanish ($\Delta S = 0$) is often taken to represent the ideal glass transition, and this is how 'liquid' and 'amorphous' states are separated in Fig. 1. Yet if the glass transition did occur when $\Delta S = 0$ during cooling, inverse melting would not be observed — in fact, the boundary between the crystal and amorphous phases would become tangential to the melting curve (Fig. 1). This is because of the freezing-in of excess entropy at the transition. Moreover, pressure-induced amorphization of ice¹⁰ and silica¹¹ appears to operate in region II, not region III. So the division between liquid and amorphous states must deviate from the $\Delta S = 0$ line.

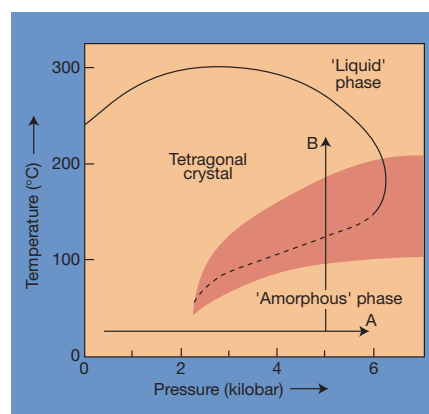


Figure 2 A simplified phase diagram of a polymeric system based on the structural² and calorimetric¹ results of Rastogi and colleagues. The transition between tetragonal and amorphous phases (dashed line) has been detected in calorimetric experiments by measurements taken under increasing pressure (arrow A) and on heating (arrow B), and it shows evidence for 'inverse melting' as predicted in Fig. 1. Studies of the transition are hampered by a hexagonal phase (shaded area).

According to Angell¹², 'strong' liquids are distinguished from 'fragile' liquids in having open network structures that are relatively insensitive to temperature change. The stronger the liquid, the further its glass transition temperature is above the $\Delta S = 0$ line¹². So it may be that conditions seen in region III, including inverse melting, may be observable only for very fragile liquids with a relatively low glass-transition temperature.

The observations of Rastogi *et al.*^{1,2} bring unaccustomed relevance to Tammann's universal, but mostly ignored, phase diagram (Fig. 1). Yet the differences between their polymeric system and others showing pressure-induced amorphization may necessitate different versions of the diagram to be drawn for differing degrees of liquid fragility. It would be timely to explore the validity of Tammann's conjectures with the greater wealth of data now at our disposal. ■

A. Lindsay Greer is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK. e-mail: alg13@cam.ac.uk

1. Rastogi, S., Höhne, G. W. H. & Keller, A. *Macromolecules* **32**, 8897–8909 (1999).
2. Rastogi, S., Newman, M. & Keller, A. *J. Polym. Sci. B* **31**, 125–139 (1993).
3. Tammann, G. *Kristallisieren und Schmelzen* 26–46 (Johann Ambrosius Barth, Leipzig, 1903).
4. Tammann, G. & Mehl, R. F. *The States of Aggregation* (Van Nostrand, New York, 1925).
5. Tammann, G. *Lehrbuch der Metallographie* 2nd edn (Leopold Voss, Leipzig, 1921).
6. Tammann, G. *A Text Book of Metallography* (Chemical Catalog Co., New York, 1925).
7. Vogel, R. *Die Heterogenen Gleichgewichte* 2nd edn (Akademische, Leipzig, 1959).
8. Baus, M. J. *J. Stat. Phys.* **48**, 1129–1146 (1987).
9. Sinkler, W., Michaelsen, C. & Bormann, R. *J. Mater. Res.* **12**, 1872–1884 (1997).
10. Mishima, O., Calvert, L. D. & Whalley, E. *Nature* **314**, 76–78 (1985).
11. Hemley, R. J., Jephcoat, A. P., Mao, H. K., Ming, L. C. & Manghnani, M. H. *Nature* **334**, 52–54 (1988).
12. Angell, C. A. *Science* **267**, 1924–1935 (1995).

Cell biology

Moving in mysterious ways

Alison M. Condliffe and Phillip T. Hawkins

In their mission to engulf and destroy invading microorganisms, certain white blood cells (leukocytes) must leave the blood stream and migrate to the sites of infection. Such cells track their targets by following increasing concentrations of chemoattractant molecules — chemotactic gradients — until at higher concentrations the engulfment response comes into operation. Failure of the chemotactic process results in increased susceptibility to infection, but excessive or inappropriate responses can lead to self-damage, as seen in a variety of inflammatory diseases.

No fewer than five papers^{1–5}, published in *Science* last month, address this vital process. Their main message is that chemotaxis requires an enzyme called phosphoinositide-3-OH kinase (PI(3)K) to generate lipid signals that result in a highly polarized signalling cascade in the cell. The upshot of the cascade is extension of lamellipodia — cell projections — at the leading edge of the cell, which is essential for directional movement (Fig. 1).

Chemoattractants are a diverse group of molecules. They include bacterial-derived peptides, for instance one known as fMLP; components of the immunological complement cascade, classically C5a; and members of the chemokine family such as interleukin-8. The leukocyte cells involved are neutrophils and macrophages, and the chemoattractants bind to receptors on their surface that are linked to the G proteins that mediate various intracellular signalling pathways.

The result is dissociation of the G protein into its α and $\beta\gamma$ subunits, and subsequent activation of targets further down the pathway.

Four isoforms of PI(3)K are known (α , β , γ and δ), but only the leukocyte-specific PI(3)K γ is activated solely by G $\beta\gamma$ subunits. PI(3)K γ had thus been tentatively implicated in translating signals from G-protein-linked receptors into specialized leukocyte responses such as chemotaxis and the respiratory burst (production of toxic oxygen-derived free radicals to kill engulfed bacteria). PI(3)K synthesizes the phosphoinositide phosphatidylinositol-3,4,5-

trisphosphate (PtdIns(3,4,5)P₃) which, together with its immediate breakdown product phosphatidylinositol-3,4-bisphosphate (PtdIns(3,4)P₂), is believed to be generated in the inner leaflet of the plasma membrane and is now accepted as a signalling molecule.

Three of the new papers^{1–3} describe the use of 'knockout' mice lacking the catalytic subunit of PI(3)K γ , p110 γ , for the first time allowing the physiological functions of PI(3)K γ to be defined. They show that under certain conditions PI(3)K γ is the sole source of PtdIns(3,4,5)P₃ in response to fMLP, C5a and interleukin-8 in leukocytes (Fig. 2), and that the downstream signalling events (including activation of protein kinase B) are dramatically reduced in the p110 γ -deficient cells.

Most notable, however, were the physiological consequences in the p110 γ -deficient animals. *In vitro*, neutrophils from these animals migrated inefficiently to a variety of chemoattractants. *In vivo*, the numbers of leukocytes migrating to sites of inflammation (for example, abdominal infection) were markedly reduced. The mice also demonstrated a reduced neutrophil oxidative burst in response to G-protein-linked activators, and other immunological defects (impaired maturation and activation of T cells). The overall result was persistence of viable bacteria in the peritoneal cavity of the knockout mice, and an impaired response to viral infection.

Chemotaxis features in many other biological processes, such as embryonic development and formation of new blood vessels, during which cells move along cytokine or growth-factor gradients. Although these molecules generally signal through tyrosine-kinase-linked receptors that are not coupled to PI(3)K γ , other PI(3)K isoforms are thought to take part in these processes. Examples include the movement of epithe-

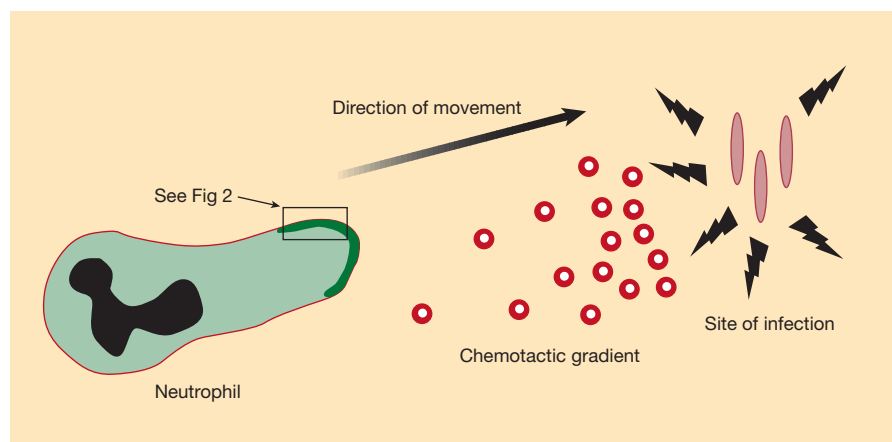


Figure 1 Cell movement in chemotaxis. Neutrophils move up a chemotactic gradient of chemoattractant generated by invading bacteria by forming lamellipodia — cell projections — at the leading edge of the cell. Chemoattractant receptors (red) are distributed uniformly over the cell surface. But the extension of lamellipodia, and thus cell movement, depends on the selective concentration of specialized proteins at the cell's leading edge.

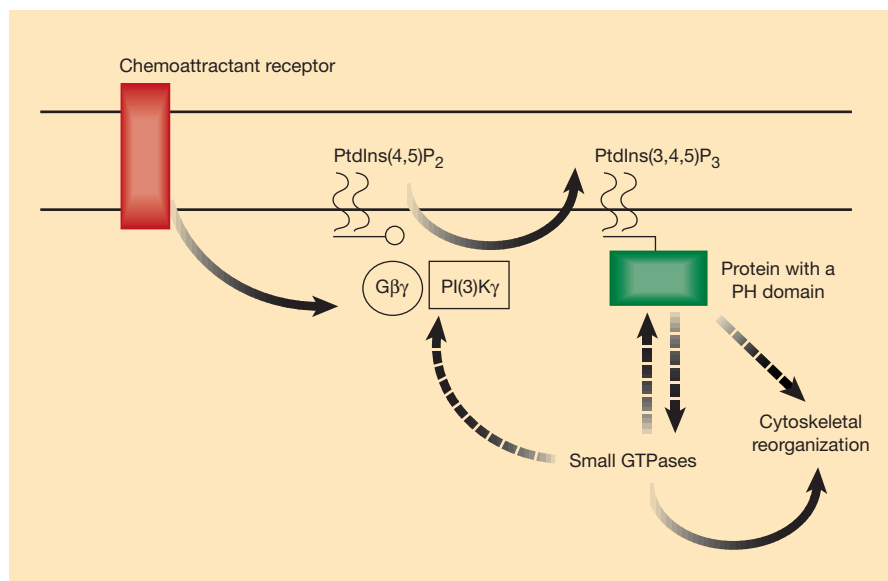


Figure 2 Signalling events in chemotaxis. On binding chemoattractant, the chemoattractant receptor signals through the G-protein $\beta\gamma$ subunit to activate PI(3)K γ . Activated PI(3)K γ generates lipid signals such as PtdIns(3,4,5)P $_3$. Proteins with pleckstrin-homology (PH) domains bind the lipid-signalling elements (perhaps under the influence of small GTPases), and activate cellular responses. Such responses include reorganization of the cytoskeleton for lamellipodial extension and cell motion.

lial cells towards platelet-derived growth factor⁶ or that of macrophage-like cells towards colony-stimulating factor-1 (ref. 7). Thus it is possible that the PI(3)K signalling pathway is generically involved in directed cell movements.

How might the PI(3)K-derived lipid signals work? Chemotaxis involves an array of responses including cytoskeletal reorganization, modulation of integrin adhesive properties and membrane trafficking. From studies of two model systems, *Dictyostelium* amoebas and neutrophil-like HL-60 cells^{4,5,8}, it seems that proteins with phosphoinositide-binding pleckstrin homology (PH) domains are selectively recruited to the leading edge of chemotaxing cells. This polarization is much sharper than the causative chemotactic gradient, and occurs despite the uniform distribution of receptors around the cell periphery. It is not clear how the steep signalling gradient is generated, but it seems likely to involve the production and degradation of phosphoinositides; in this context it is interesting that mice deficient in SHIP (SH2-containing inositol-5-phosphatase), an enzyme that hydrolyses PtdIns(3,4,5)P $_3$, suffer from lethal infiltration of the lungs by macrophages and neutrophils⁹. So abnormal persistence of the PtdIns(3,4,5)P $_3$ signal might lead to excessive inflammation.

Another of the new papers⁵ suggests that members of the Rho family of small GTPases, GTP-regulated molecular switches, are involved in the differential recruitment of PH-domain-containing proteins to the cell's leading edge. The Rho GTPases have been implicated in other cell-signalling systems both upstream and downstream of PI(3)Ks,

and mediate the polymerization of actin (the main component of the cytoskeleton). So they might well have a key part in the cooperative signalling events leading to chemotaxis. This possibility is supported by the attenuated chemotactic behaviour of cells derived

from mice in which Rac-2, a member of the Rho family, was knocked out¹⁰.

Clearly, there is still a long way to go before we have a comprehensive description of chemotaxis. The precise means by which asymmetrical signalling is generated remain unclear. The identity of the lipid signals and indeed their cellular targets have not been convincingly defined, although preliminary evidence implicates protein kinase B (ref. 10) and Bruton's tyrosine kinase¹¹, both of which have PH domains. Finally, there is the need to unravel the relationship between this signalling system and others involved in coordinating chemotactic responses, such as those using cyclic adenosine monophosphate and intracellular calcium.

Alison M. Condliffe is in the Department of Medicine, University of Cambridge Clinical Medical School, Addenbrookes and Papworth Hospitals, Hills Road, Cambridge CB2 2QQ, UK.

Phillip T. Hawkins is at The Babraham Institute, Babraham, Cambridge CB2 4AT, UK.

e-mail: phillip.hawkins@bbsrc.ac.uk

1. Sasaki, T. *et al.* *Science* **287**, 1040–1046 (2000).
2. Li, Z. *et al.* *Science* **287**, 1046–1049 (2000).
3. Hirsch, E. *et al.* *Science* **287**, 1049–1053 (2000).
4. Jin, T., Zhang, N., Long, Y., Parent, C. A. & Devreotes, P. N. *Science* **287**, 1034–1036 (2000).
5. Servant, G. *et al.* *Science* **287**, 1037–1040 (2000).
6. Wennström, S. *et al.* *Oncogene* **9**, 651–660 (1994).
7. Vanhaesebroeck, B. *et al.* *Nature Cell Biol.* **1**, 69–71 (1999).
8. Meili, R. *et al.* *EMBO J.* **18**, 2092–2105 (1999).
9. Helgason, C. D. *et al.* *Genes Dev.* **12**, 1610–1620 (1998).
10. Roberts, A. W. *et al.* *Immunity* **10**, 183–196 (1999).
11. Nore, B. F. *et al.* *Eur. J. Immunol.* **30**, 145–154 (2000).

Molecular electronics

Pushing electrons around

Mark Ratner

Most of the designs and prototypes for 'molecular electronics' circuits and devices¹ involve an interface between a molecule and a conductive (metal or semiconductor) bulk material. At this interface the discrete electronic structure of the molecule confronts the continuous electronic band structure of the bulk material, or electrode. The interaction between these structures results in a number of striking phenomena, including charge transfer. But building these devices is not easy; there are problems with the energy threshold and stability of currents through the molecules.

On page 166 of this issue Vilan *et al.*² show how molecular layers (rather than individual molecules) can control the electrical behaviour of a 'classical' metal-semiconductor device. This work represents a new approach to the use of molecular layers in electronics because the control is achieved without any electron transport through the molecules themselves.

The device that Vilan *et al.*² modify is a

tunnel diode — this is a type of diode that relies on electron tunnelling, whereby electrons make use of quantum mechanics to tunnel through a potential barrier that they can't cross classically. (A typical semiconductor diode is a junction through which electrons can only flow in one direction.) In a tunnel diode the electrons can tunnel or pass across the junction only when the voltage is positive. Vilan *et al.* modified their metal-semiconductor diode by adsorbing tartaric acid molecules on to the semiconductor crystal (gallium arsenide, GaAs). The molecules bind very strongly to the crystal surface, producing a single layer on which the metal contact (in this case gold) was gently deposited. A benzene group was also attached to the tartaric acid molecules to which groups with various dipole moments could be added. The authors found that by switching these groups, and therefore changing the dipole moment, they could alter the current flowing through the diode (Fig. 1, overleaf).

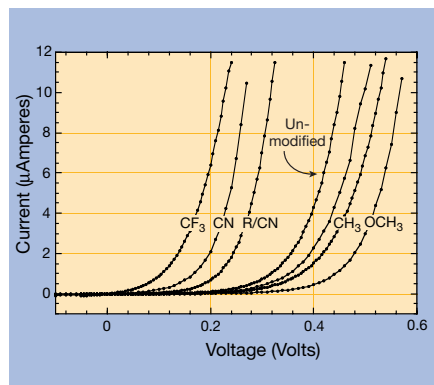


Figure 1 Current-voltage plot for the 'molecular electronics' device built by Vilan *et al.*². A characteristic diode junction allows current to flow only when positive voltage is applied. The curves shown here are recorded from metal-semiconductor junctions that are identical except for the group attached to a molecular layer on the semiconductor surface. The groups have either increased or decreased dipole moments compared with the unmodified junction, and the change in current flow through the modified junctions is due to this dipole control.

Interfaces between electrodes and individual molecules are required in 'molecular wire' circuits, whose preparation, measurement and modelling are being intensely studied^{1,3-11}. The energy properties that dominate such circuits are the effective work function, $q\phi$, of the metal and the electron affinity, χ_{sc} , of the semiconductor, both of which characterize the transport of charges in these materials. A molecular layer can produce a substantial shift in the values of these parameters from $q\phi^0$ or χ_{sc}^0 of the bare electrode. This modification can be understood in several ways, ranging from simple electrostatics to reasonable electronic structure calculations¹². Qualitatively, much of the change from $q\phi^0$ or χ_{sc}^0 to $q\phi$ or χ_{sc} arises from the way in which the molecular dipole distribution at the interface perturbs the potential felt by the electrons in the bulk phase. This perturbation will substantially shift the current-voltage characteristics of a molecular wire circuit^{1,3,4} or the voltage dependence of the charge injection in organic light-emitting diodes.

Because the molecular layers can so dramatically alter the surface work function, they can provide a substantial level of control over the electrode properties. Vilan *et al.*² show that the tunnelling barrier height for a pure metal-semiconductor junction can be controlled by the molecular dipole layer, so that the net barrier height $q\phi^b$ for the junction is given by $q\phi^b = q\phi_{metal}^0 - \chi_{sc} + q\phi_{dipole}$. The unmodified junction height (without the molecular layer) is given by just the metal and semiconductor terms, as defined previously. The extra term, $q\phi_{dipole}$, is the change in barrier height due to the molecular layer. By

changing the molecular dipole moment value and sign, the barrier height can be altered.

Vilan *et al.* demonstrate this control by changing the polarity of the group attached to the molecular layer — from dipoles, such as CF_3 and CN , that increase $q\phi_{dipole}$, to others (CH_3 and OCH_3) that reduce it. As the dipole effect changes, so does the net barrier height $q\phi^b$. The actual barrier heights were calculated from the current-voltage characteristics of the junctions (Fig. 1), and the value increases linearly by 90 millielectron volts on changing from CN to OCH_3 (see their Fig. 2 on page 167). What is so striking is that electron transport through the molecules seems to contribute little or nothing to the change in junction properties — the effect is due entirely to the molecular dipoles.

Several recent contributions have shown that molecular layers can modify the surface properties of semiconductors in a significant way¹³⁻¹⁵. The demonstration by Vilan *et al.* that changes in the molecular dipole moment are linearly reflected in changes in the junction barrier is an important advance in the engineering of charge transport at molecular electronics interfaces. In cases where the current is actually carried through the molecule, the change in charge injection properties can alter the conductance by changing the effective state density. This means that states into

which the electrons tunnel become available at altered energies. If the charge is transported through the films by gaps, shorts or direct tunnelling (that is, if the monolayer is not perfect), tuning the molecular dipole layer can, as Vilan *et al.* show, lower or raise the threshold for tunnelling conductivity. Both molecular wire circuits and charge injection in organic devices vary strongly with effective barrier heights; the adsorption of molecular dipole layers apparently offers a general and practical approach to customizing these barriers.

Mark Ratner is in the Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3113, USA.

e-mail: ratner@chem.nwu.edu

1. Jortner, J. & Ratner, M. A. (eds) *Molecular Electronics* (Blackwell, London, 1997).
2. Vilan, A., Shanzer, A. & Cahen, D. *Nature* **404**, 166-168 (2000).
3. Tian, W. D. *et al.* *J. Chem. Phys.* **109**, 2874-2882 (1998).
4. Reed, M. A. *et al.* *Science* **278**, 252-254 (1997).
5. Dai, H. *et al.* *Science* **272**, 523-526 (1996).
6. Joachim, C. & Vinuesa, J. F. *Europhys. Lett.* **33**, 635-640 (1996).
7. Dekker, C. *Phys. Today* 22 May (1999).
8. Bumm, L. A. *et al.* *J. Phys. Chem. B* **103**, 8122-8127 (1999).
9. Hall, L. E., Reimers, J. & Hush, N. S. *J. Phys. Chem.* **112**, 1510-1521 (2000).
10. Mujica, V. *et al.* *J. Phys. Chem.* (in the press).
11. Emberly, E. & Kirczenow, G. *Nanotechnology* **10**, 285-289 (1999).
12. Vondrak, T., Cramer, C. J. & Zhu, X. *J. Phys. Chem. B* **103**, 8915-8919 (1999).
13. Campbell, L. H. *et al.* *Appl. Phys. Lett.* **71**, 3528-3530 (1997).
14. Kruger, J. *et al.* *Adv. Materials* (in the press).
15. Cohen, R. *et al.* *J. Am. Chem. Soc.* **121**, 10545-10551 (1999).

Entomology

Love is not puffed up

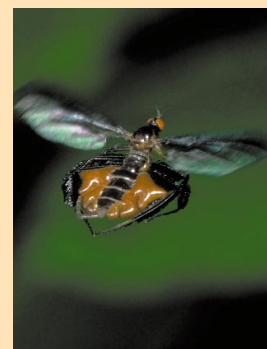
When it comes to picking a mating partner, the females of most species usually get to do the choosing. Sometimes, as in the case of the long-tailed dance fly *Rhamphomyia longicauda*, the roles are reversed. David Funk and Douglas Tallamy show in *Animal Behaviour* (**59**, 411-421; 2000), however, that even here the females seem to be holding all the cards.

Female dance flies cannot hunt for prey for themselves, and instead rely on nuptial gifts presented to them by the males in exchange for mating. This trade-off takes place in female leks, where females gather and wait for the males to bring them food. In contrast to many species, this means that it is the males that get to do the picking. But Funk and Tallamy reveal that the females have evolved a clever way to increase their odds of being chosen.

The males prefer to mate

with females with swollen abdomens. The likely reason is that abdomen size is perceived to be a good indicator of egg maturity — for a related species, *R. sociabilis*, multiple regression analysis revealed a significant relationship between these two variables. Funk and Tallamy propose that, given that males want their sperm to be the ones that fertilize the egg, waiting until the eggs are mature is a good way to ensure that this happens.

But female *R. longicauda* cheat. The females puff up their abdomens (pictured) — probably by 'swallowing' air — so creating the impression that their eggs are more mature than they really are. (Multiple regression analysis showed no significant relationship between egg maturity and the size of the inflated abdomen.) But it seems that the males of this species are still fooled. The



males hover beneath the females, and all that they see is the expanded shape of the abdomen. They have no further clues to the actual state of the females' eggs.

All of this means that females whose eggs are not yet mature can gain the sustenance needed to complete egg development at least once. It seems that, contrary to appearances, it is the females who still retain the upper hand in this mating game.

Amanda Tromans

Plate tectonics

The Antarctic connection

Richard G. Gordon

Today, Antarctica seems to be part of a single tectonic plate. But there is clear palaeomagnetic evidence that, in the past, there has been a large relative rotation between different parts of the continent¹. A central issue has been how much, if any, of this motion has occurred during the past 80 million years or so. Cande *et al.*² (page 145 of this issue) now present data that reveal an extinct plate-tectonic boundary between the two parts of Antarctica, East and West, thereby enabling them to estimate that motion. East Antarctica is the larger and has been a stable part of a nearly rigid plate for hundreds of millions of years. West Antarctica, by contrast, is probably an amalgamation of many smaller pieces that were assembled in the past few hundred million years. The significance of Cande and colleagues' study goes well beyond Antarctica, however, because motion in Antarctica is a central element in reconstructing plate-tectonic motions around the globe.

To understand the evolution of the plate margins surrounding the Pacific Ocean, one would like to know the displacement history of the plates of the Pacific basin relative to the surrounding continental plates. During seafloor spreading the polarity of the geomagnetic field, which reverses on average every half million years or so, is locked into newly created and cooling crust, thus preserving a record of the history of relative motion of the sea floor flanking a mid-ocean ridge. Processes at converging plate boundaries, however, leave no useful record for estimating the relative plate motions that have occurred across them. To estimate the motion across such boundaries, a so-called 'plate-motion circuit' through a series of mid-ocean ridges is needed.

The present circum-Pacific margin has only one very small segment of mid-ocean ridge, in the Gulf of California along the coast of Mexico, and it has existed for only a few million years. So, to estimate the motion between Pacific-basin plates and the surrounding continental plates for any time except the past few million years, a plate-motion circuit must be constructed through mid-ocean ridges³.

Figure 1 illustrates a set of such plate-motion circuits. It shows the central role of the Antarctic plate, which is nearly surrounded by mid-ocean ridges. To estimate the motion of the Pacific plate relative to the North American plate, for example, the relevant circuit is the Pacific plate to the Antarctic plate (via the Pacific–Antarctic Rise) to the African plate (via the Southwest Indian

Ridge) to the North American plate (via the Mid-Atlantic Ridge), with the relative motion between each of these plate pairs being summed. Similarly, the circuit must pass through Antarctica to estimate the motion between any Pacific-basin plate and any of the surrounding continental plates — North America, South America, Eurasia or Australia. These plate-motion circuits can be used, at most, for the past 80 million years, the time during which the Pacific–Antarctic Rise has existed.

These same plate-motion circuits are the main yardsticks that one has for estimating the relative motion between the hotspots in the Pacific Ocean and those in the Atlantic and Indian oceans. Hotspots are long-lived sources of volcanism, such as that which is now active at Kilauea volcano in the Hawaiian islands. Hotspots are thought to be the surface manifestations of plumes rising from deep in the mantle. As a plate moves relative to a plume beneath it, a trail of extinct volcanoes, such as the Hawaiian–Emperor island and seamount chain, is created.

The original plate reconstructions for the south Pacific, Antarctica, Australia and New Zealand, published in 1975, implied that about 500 km of motion was required between East and West Antarctica since about 80 million years ago⁴. Estimates of the past motions of Pacific-basin plates relative

to the circum-Pacific plates showed that these reconstructions are extremely sensitive to assumed motion between East and West Antarctica⁵.

Furthermore, two approaches for estimating the northward motion of the Pacific plate relative to Earth's spin axis have produced estimates that disagree; motion between East and West Antarctica might partly explain this disagreement⁶. In the first approach, the northward motion is estimated directly from Pacific plate palaeomagnetic data, which record the ancient time-averaged direction of Earth's magnetic field. The time-averaged inclination, which is the angle that the magnetic field makes with the horizontal, varies in a predictable way with palaeolatitude, from which northward motion can be inferred. In the second approach, the amount of northward motion of the Pacific plate is inferred from palaeomagnetic data reconstructed from other plates through the global circuit through Antarctica, assuming that Antarctica is a single plate⁶.

There also has long been a large misfit between observed and hypothetical Pacific hotspots tracks predicted in a reference frame fixed to Atlantic and Indian Ocean hotspots. One explanation⁷ for the misfit is substantial motion between Pacific and non-Pacific hotspots. A second explanation^{8,9} is that the hotspots in the Pacific are fixed relative to non-Pacific hotspots, and that motion between East Antarctica and West Antarctica causes the misfit.

Since 1975, new data and interpretations have reduced the size of the plate-reconstruction misfits that required motion

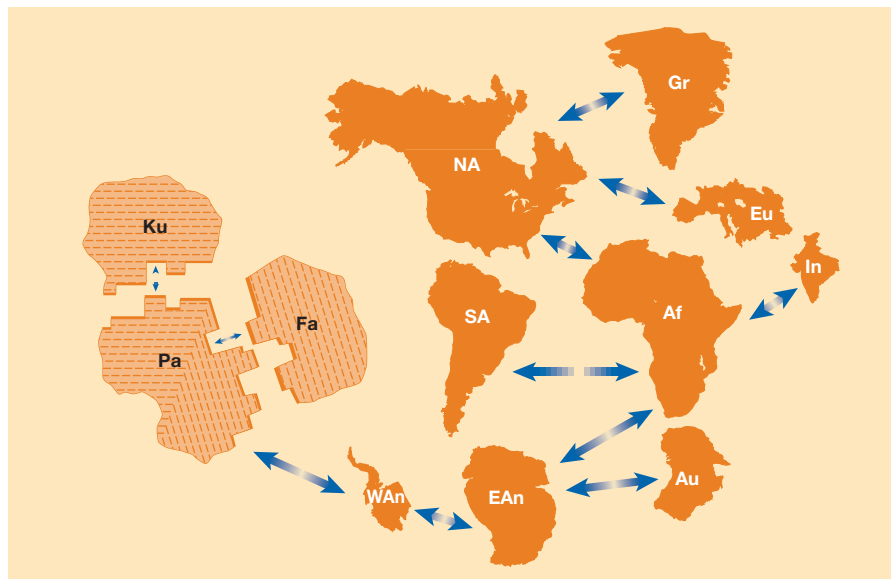


Figure 1 Global plate-motion circuits for much of the past 80 million years. The double-headed arrows point to pairs of plates separated by a mid-ocean ridge from which the relative plate motion of the plate pair can be determined. For example, the motion of the Kula plate relative to the North American plate is determined by the plate-motion circuit Kula (Ku) to Pacific (Pa) to West Antarctic (WAn) to East Antarctic (EAn) to Africa (Af) to North America (NA). Au, Australia; Eu, Eurasia; Fa, Farallon; Gr, Greenland; In, India; SA, South America. (Redrawn from refs 6, 12.)

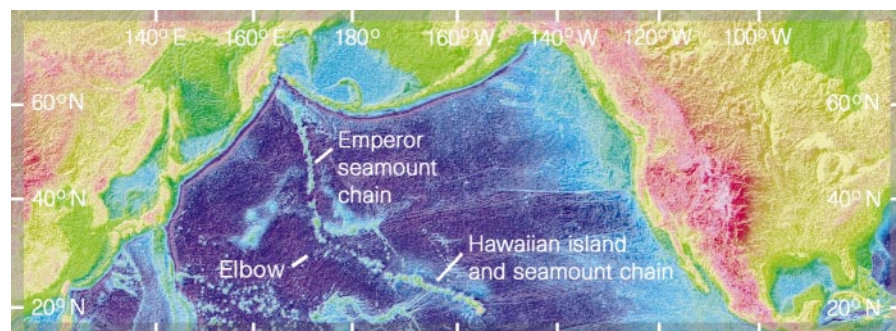


Figure 2 Topography of the north Pacific sea floor, showing the 'elbow' where the Hawaiian island and seamount chain meets the Emperor seamount chain. (Map reproduced from ref. 13, courtesy of Walter H. F. Smith.)

between East and West Antarctica. Some workers have gone so far as to assert that there has been no such motion during the past 80 million years. Much of the ambiguity has been due to a lack of essential geophysical data from high-latitude regions, which are difficult to survey because of the presence of sea ice.

Cande *et al.*² have produced such data from marine geophysical surveys north of the Ross Sea and in the South Tasman Sea (see Figs 1 and 2 of the paper on pages 145 and 146). These data allow them to directly estimate motion between East and West Antarctica. They have discovered an extinct spreading centre north of the western Ross Sea embayment, which is flanked by a set of linear magnetic anomalies that formed between 43 and 26 million years ago. These anomalies indicate a spreading rate of some 12 mm per year, about the same rate as the slowest rates of seafloor spreading observed today.

Cande *et al.* combine these and other data to show that there was about 180 km of separation in the western Ross Sea embayment, and hence 180 km of motion between East and West Antarctica, between 43 and 26 million years ago. This work not only places firm bounds on how much motion could have occurred between East and West Antarctica; it also — for the first time — permits rigorous estimates of this motion to be incorporated into the global plate-motion circuit and into estimates of the motion of Pacific hotspots relative to non-Pacific hotspots.

The results bear directly on the question of the origin of the 'elbow' in the Hawaiian–Emperor chain in the north Pacific (Fig 2). Ages along the chain, which record the motion of the Pacific plate relative to the Hawaiian hotspot, increase monotonically to the northwest from zero (that is, active volcanism) at its southeast end (for example Kilauea on the big island of Hawaii) to about 80 million years near where it intersects the Aleutian trench. The chain consists mainly of two nearly straight segments, the older north-northwest-striking Emperor segment with ages from about 80 to 43 million years, and the younger west-northwest-striking

Hawaiian segment with ages less than 43 million years. These two straight segments meet at an angle of about 120° in an 'elbow', which is widely, but not universally, interpreted as recording a 60° change at 43 million years ago in the direction of Pacific plate motion relative to the Hawaiian and other Pacific plumes.

Reconstructions¹⁰ of the Pacific plate relative to Atlantic and Indian Ocean hotspots that use Cande and colleagues' estimates indicate no sharp change in motion of the Pacific plate relative to the Atlantic and Indian Ocean hotspots at about 43 million years ago. If so, it implies that there was a sharp and sudden change in the motion of Pacific hotspots relative to those in other ocean basins, which is difficult to understand in view of what is known about convection in Earth's mantle¹¹.

In any event, the new work of Cande *et al.* constitutes great progress. It will have a pivotal influence on the future course of research on global tectonics.

Richard G. Gordon is at the Department of Geology and Geophysics, University of California, Berkeley, California 94720-4767, USA, on sabbatical leave from the Department of Geology and Geophysics, Rice University.

e-mail: rgg@rice.edu

1. Luyendyk, B. *et al.* *Tectonics* **15**, 122–141 (1996).
2. Cande, S. C., Stock, J. M., Müller, D. & Ishihara, T. *Nature* **404**, 145–150 (2000).
3. Atwater, T. & Molnar, P. in *Proceedings of Conference on Tectonic Problems of San Andreas Fault System* (eds Kovach, R. L. & Nur, A.) 136–148 (Stanford Univ. Publ. Geol. Sci., 1973).
4. Molnar, P., Atwater, T., Mammertickx, J. & Smith, S. M. *Geophys. J. R. Astron. Soc.* **40**, 383–420 (1975).
5. Jurdy, D. M. *J. Geophys. Res.* **84**, 6796–6802 (1979).
6. Acton, G. D. & Gordon, R. G. *Science* **263**, 1246–1254 (1994).
7. Molnar, P. & Stock, J. *Nature* **327**, 587–591 (1987).
8. Morgan, W. J. in *Oceanic Lithosphere* (The Sea, Vol. 7) (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
9. Duncan, R. A. *Tectonophysics* **74**, 29–42 (1981).
10. Raymond, C. A., Stock, J. M. & Cande, S. C. in *History and Dynamics of Global Plate Motions* (eds Richards, M., Gordon, R. G. & van der Hilst, R. D.) (Am. Geophys. Un., Washington DC, in the press).
11. Bercovici, D., Ricard, Y. & Richards, M. in *History and Dynamics of Global Plate Motions* (eds Richards, M., Gordon, R. G. & van der Hilst, R. D.) (Am. Geophys. Un., Washington DC, in the press).
12. Engebretson, D. C., Cox, A. & Gordon, R. G. *Geol. Soc. Am. Spec. Pap.* **206** (1985).
13. http://www.ngdc.noaa.gov/mgg/announcements/images_predict.HTML

Daedalus

Expiry date guaranteed

Last week Daedalus invented the short-life pet. By repeatedly dividing a newly fertilized ovum in vitro, he obtained large numbers of single-cell ova, each of which had already undergone many cell divisions. Every time a cell divides, each of its chromosomes loses a telomere unit from its termination; when they have all gone, it can divide no longer, and is doomed. So Daedalus's telomere-depleted ova have a limited life ahead of them. Brought to term in a surrogate mother, they are sure to die young — making them perfect pets.

This process has further implications. Do even human identical twins, who have undergone one more cell division than the rest of us, have a slightly reduced lifespan? Probably not. Few of us die from telomere exhaustion; indeed, even Daedalus finds it hard to guess how it would show itself. With luck it will be a mercifully swift extinction, as one whole class of body cells — in the gut, say, or the liver — suddenly ceases to replicate. If so, Daedalus sees applications in humane animal husbandry. At present, sadly, all farmed animals have to be deliberately killed. Any animal that dies a 'natural death' has probably succumbed to some pathogen, and its meat is unsafe to eat. But an animal that dies of telomere exhaustion would be entirely safe. With proper timing, it could be arranged to drop dead just after having reached its ideal size and condition.

Even better, repeated ovum division creates unlimited numbers of such animals, all identical. Implanted into surrogate mothers, they would all have the same predictable lifespan. DREADCO agronomists are now studying the logistics of a humane agriculture in which identical ova are implanted at regular intervals into a production line of surrogate mothers. The animals are brought to term and allowed to grow up normally. On the predicted day of their death, they are transported to the slaughterhouse by modern 'just in time' delivery methods. They walk onto the conveyor belt in the correct order, and each drops dead just at it reaches the butchery section. No violence is necessary. Animal-rights enthusiasts will cheer, consumers will welcome the reliable consistency of the products, and accountants will revel in the utter predictability of it all. **David Jones**

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Urban benzene and population exposure

People aren't just at risk from carcinogenic benzene when they are out on city streets.

Benzene pollution emanating from motor traffic can cause leukaemia^{1–3}, with the risk being estimated at about four cases per million among people who experience lifelong exposure to benzene concentrations of $1 \mu\text{g m}^{-3}$ in air^{4–6}. But we show here that personal exposure, and therefore risk estimates, cannot simply be estimated from environmental concentrations of benzene. Using a new sampling device that monitors both of these parameters, we have discovered that people living in different European cities are exposed to concentrations of benzene that may be twice as high as the urban average.

We have monitored benzene pollution in each of the European towns of Antwerp, Athens, Copenhagen, Murcia, Padua and Rouen at 100 sampling sites distributed over a multiscale grid drawn on the town map⁷. Every two months (from September 1997 to September 1998), these sites were tested from Monday morning to Friday afternoon by using a radial symmetry passive sampler device (termed a *radiello*⁸).

This device relies on the diffusion of gas molecules as a result of a concentration gradient across a diffusive barrier: diffusing molecules are captured by an adsorbing material that has a constant benzene-uptake rate of $80 \times 10^{-6} \mu\text{g min}^{-1}$ for each $1 \mu\text{g m}^{-3}$ present in air. Concentrations are calculated from the amount collected and the exposure time. Having tested the sampler reliability both in a standard atmosphere chamber and at the actual sampling sites, we found a maximum bias value of 4.45% and a coefficient of variation of 2.5–22.0% for 120 samplers exposed for 4.5 days to benzene concentrations of $1.5\text{--}47 \mu\text{g m}^{-3}$.

We monitored 50 volunteers and their homes in each town. These volunteers were non-smokers and were equally represented by people exposed to traffic fumes as part of their jobs (police, postal workers, street sweepers, stall-holders, and bus and taxi drivers) and by non-exposed people (students, teachers and clerks), who all wore the sampler from Monday morning to Friday afternoon.

The results (Fig. 1) were derived from 6,205 measurements, of which 50.7, 25.1 and 24.2% represent environmental, personal and home exposure records, respectively. The urban pollution level, presented as the annual average, is seen to increase in European towns from north to south (Fig. 1a): this may be explained by a difference in prevailing meteorological conditions, such as local wind speeds (Fig. 1b).

However, personal exposure data and measurements taken inside homes do not reflect this difference in urban pollution between northern and southern European towns: except in Athens, the population exposure exceeds the average urban concentration and is higher indoors than outdoors (Fig. 1c).

This surprising finding may be explained by the hourly benzene concentration oscillating between very low values during the night and very high values during the middle of the day and the evening.

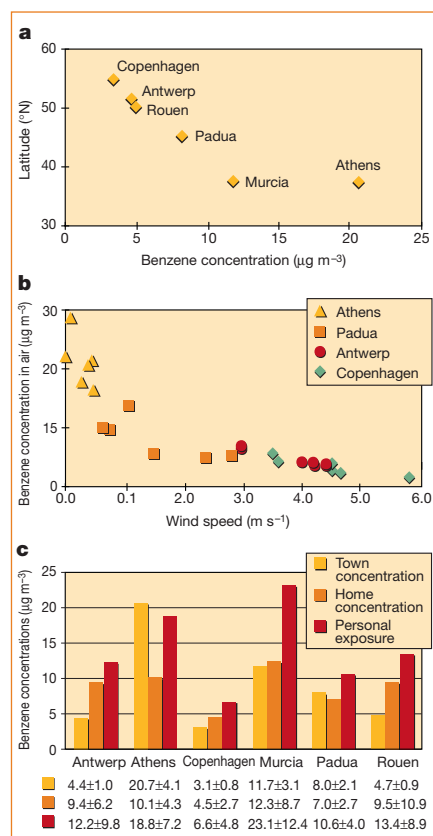


Figure 1 Benzene personal exposure level cannot be calculated as a time-weighted average of outdoor and indoor mean concentrations owing to large hourly urban pollution oscillations. **a**, Benzene urban pollution seems to increase from northern to southern European towns. **b**, This observation may be explained by differences in wind speeds, and therefore ventilation, clearing benzene pollution at different rates in northern and southern towns. In estimating the local population exposure, the reduction in the pollution of northern compared with southern towns is balanced by an opposite trend in indoor pollution. **c**, Population exposure depends on both outdoor and indoor pollution. Measurements using the *radiello* device show that personal exposure is, on average, twice the mean urban pollution level. Annual values for benzene concentrations, presented as averages from six monitoring campaigns, in the street and home and the resulting personal exposure are shown. Values are average concentrations in $\mu\text{g m}^{-3}$, with 95% confidence intervals.

As most people are about in the streets when the benzene concentration is higher than the daily average, then their actual exposure when out of doors is higher than the value calculated from the daily urban average benzene concentration and the time spent outdoors.

But this outdoor exposure represents only a fraction of the total: as people spend on average 59.1% of their day at home, the contribution from indoor exposure is important. In the European towns studied here, the average pollution level at home turns out to be 1.51 times the outdoor street level, with the exception of Athens, where volunteers live far from the town centre.

The value of the domestic-to-urban pollution ratio rises from southern to northern Europe. This tendency evens out the difference in the level of personal exposure and the level of urban pollution: when no longer exposed out of doors, people are subject to indoor pollution, which is worse in Antwerp, Rouen and Copenhagen, comparable in Padua and Murcia, and better in Athens.

To explain these findings, we suggest that pollution indoors is caused by benzene entering from the streets outside, as shown by the good fit of respective seasonal trends (data not shown). The pollution indoors is generally higher than outdoors, possibly because of an imbalance between the flow of pollutant from outside and its removal from inside to outside. In other words, the house itself could be acting as a flywheel created by adsorbent surfaces on walls, floors and furnishings. This idea is supported by the lower indoor pollution in southern European towns: in northern European houses, carpets, linoleum and wood surfaces are favoured, whereas tiling, marble and bare walls are typically used in southern Europe.

Vincenzo Cocheo*, Paolo Sacco*, Caterina Boaretto*, Emile De Saeger†, Pascual Perez Ballesta†, Henrik Skov‡, Eddy Goelen§, Norbert Gonzalez||, Antonia Baeza Caracena¶

*Fondazione Salvatore Maugeri-IRCCS, via Svizzera 16, 35127 Padova, Italy
e-mail: fsmgd@tin.it

†Joint Research Centre, 21020 Ispra, Italy

‡National Environmental Research Institute, Frederiksborgvej 399, 4000 Roskilde, Denmark
§Vlaamse Instelling voor Technologisch Onderzoek, Boeretang 200, 2400 Mol, Belgium

||Institut National de l'Environnement Industriel et des Risques, 60550 Verneuil-en-Halatte, France

¶Departamento de Ingeniería Química-Murcia, Universidad de Murcia, 30071 Murcia, Spain

1. Snyder, R., Witz, G. & Goldstein, B. D. *Environ. Health Perspect.* **100**, 293–306 (1993).
2. Ward, J. B. *et al. Mutat. Res.* **268**, 49–57 (1992).
3. Vigliani, E. C. & Saita, G. N. *Engl. J. Med.* **271**, 872–876 (1964).
4. Rinsky, R. A. *et al. Am. J. Ind. Med.* **2**, 217–245 (1981).
5. Bond, G. G., McLaren, E. A., Baldwin, C. L. & Cook, R. R. *Br. J. Ind. Med.* **43**, 685–691 (1986).
6. World Health Organization *Updating and Revision of the Air Quality Guidelines for Europe* Report no. EUR/ICP/EHAZ9405/MT 12 (WHO Regional Office for Europe, Brussels, 1996).
7. Kumar, N. & Russel, A. G. *Atmos. Environ.* **30**, 1099–1116 (1996).
8. Cocheo, V., Boaretto, C. & Sacco, P. *Am. Ind. Hyg. Assoc. J.* **57**, 897–904 (1996).

Fisheries

Climate variability and North Sea cod

The stock of North Sea cod is under pressure because of overfishing, and we show here that it is also threatened by a decline in the production of young cod that has paralleled warming of the North Sea over the past ten years. The combination of a diminished stock and the possible persistence of adverse warm conditions is endangering the long-term sustainability of cod in the North Sea. To decrease the risk of collapse, fishing pressure must be reduced.

Over the past four decades, cod in the North Sea has been a valuable fishery, yielding an average of 200,000 tonnes per year. Nowadays, the catch is mainly constituted of fish that are younger than three years old, most of which are immature. This resource is currently managed by imposing landing quotas based on annual stock assessment:

the latest advice from the International Council for the Exploration of the Sea (ICES) is to reduce catches by 40–60%.

For North Sea cod, apart from fish spawned in 1996 (termed the 1996 year-class), the annual number of one-year-old cod in the population (known as the recruitment) has been at or below the long-term average for over a decade¹. The recruitment of cod spawned in 1997 was the lowest for 30 years. Taking account of the dependence of recruitment on stock size alone, it has been proposed² that the collapse of this stock may be imminent, as the spawning-stock biomass is currently at a low level.

There is evidence, however, that for many fish stocks there is an environmentally driven variability in recruitment³. The recruitment of cod in the North Atlantic appears to be related to sea temperature for stocks located at the latitudinal limits of the species' distribution⁴. In the Northern Hemisphere, increasing temperatures are favourable for stocks at the highest latitudes but detrimental for those at the southern limit⁵. Cod in the North Sea are near the southern boundary of their range and, historically, strong year-classes have been associated with lower-than-average temperatures during the first half of the year (Fig. 1a).

Weak year-classes have also occurred during cold years (such as 1986 and 1987), but only when the spawning-stock biomass was low. So when the spawning-stock biomass is low, as now, the likelihood of good recruitment is low and the effect of temperature is less pronounced (Fig. 1a). A change in temperature patterns might prevent the stock from being able to produce recruitments as high as those that occurred during the 1960s and 1970s, even if the spawning-stock biomass were to rebuild to the abundant levels of that period.

Since 1988, mean sea temperatures during the first half of the year in the North Sea have been higher than during the previous three decades (Fig. 1b). During this period, annual recruitment levels have been low, with the exception of 1996, when cold conditions prevailed. Since 1997, warmer conditions have returned to the North Sea and the 1997 and 1998 year-classes have coincidentally been the poorest on record¹.

The North Sea cod stock is currently dominated by immature fish of less than five years old, and both the stock and fishery are dependent upon years of strong

recruitment. The 1996 year-class was the strongest for over a decade but, at the present high exploitation rate of young cod, few individuals have survived to reach sexual maturity.

The combination of this exploitation with the recent changes in North Sea temperature, a low spawning-stock biomass and a stock dominated by young immature individuals, means that fishery managers must take precautionary measures⁶. In order to give the mature stock a chance to rebuild, fishing mortality rates need to be reduced to at least the precautionary levels advised by ICES¹.

Carl M. O'Brien, Clive J. Fox,
Benjamin Planque, John Casey

Centre for Environment, Fisheries and Aquaculture
Science, Lowestoft Laboratory, Pakefield Road,
Lowestoft NR33 0HT, UK
e-mail: c.m.obrien@cefas.co.uk

1. ICES Cooperative Research Report Report of the ICES Advisory Committee on Fishery Management (ICES, Copenhagen, 1999); on <http://www.ices.dk/>
2. Cook, R. M., Sinclair, A. & Stefánsson, G. *Nature* **385**, 521–522 (1997).
3. Cushing, D. H. *Excellence in Ecology 7* (Ecology Institute, Oldendorf/Luhe, Germany, 1996).
4. Planque, B. & Frédou, T. *Can. J. Fish. Aquat. Sci.* **56**, 2069–2077 (1999).
5. Coutant, C. C. *Environ. Biol. Fish.* **18**, 161–172 (1987).
6. Longhurst, A. *Fish. Res.* **38**, 101–108 (1998).

Marine ecology

Bleaching patterns in reef corals

Coral reefs are under threat from the effects of bleaching, in which symbiotic algae or their photosynthetic pigments are destroyed by increased sea temperatures and solar radiation^{1,2}. Here we show that the bleaching susceptibility of *Goniastrea aspera*, a shallow-water Indo-Pacific coral, can be predicted from its history of exposure to solar radiation, demonstrating how experience can shape coral bleaching patterns.

At our study site, Phuket in Thailand, sea temperatures are maximal in May³. In 1991 and 1995 sea temperatures were anomalously high: on both occasions many coral species suffered temperature-induced bleaching, but *G. aspera* colonies were only bleached on their east-facing surfaces. *G. aspera* at Phuket also suffers annual solar bleaching, which is restricted to the west-facing surface, during January–March⁴. Solar bleaching is a photochemical effect caused by photosynthetically active radiation (PAR, of wavelengths 400–700 nm), rather than by ultraviolet radiation or solar heating effects^{4,5}.

We investigated three possible mechanisms to explain the greater susceptibility of the east rather than the west surface of

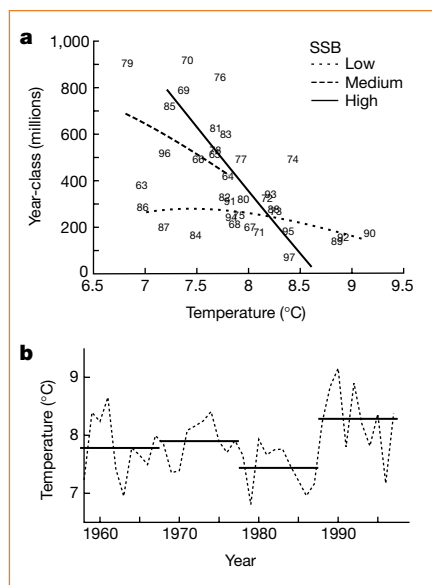


Figure 1 Cod recruitment and temperature in the North Sea. **a**, Year-class strength (millions of one-year-old fish) plotted against temperature in the year spawned. The three lines show the relationship between recruitment and temperature at three levels of spawning-stock biomass (SSB) (low: 80,000 tonnes; medium: 160,000 tonnes; high: 240,000 tonnes) based upon a three-dimensional spline fit. Recruitment and SSB are from ref. 1. **b**, Seasonal sea-surface temperature averaged over February to June (dashed line). Horizontal lines indicate decadal average. Temperatures are derived from the Comprehensive Ocean Atmosphere Dataset and provided by the National Center for Atmospheric Research (Boulder, Colorado).

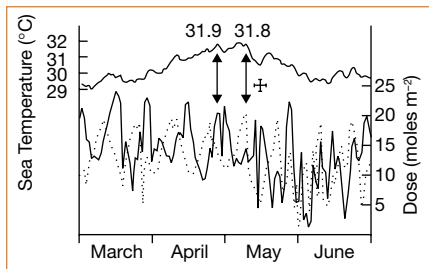


Figure 1 Daily dose of photosynthetically active radiation (PAR) underwater at the depth of *G. aspera* colonies in 1995. Solid line, west surface; dotted line, east surface. Top trace, daily mean sea temperature; double-headed arrows, maximum combined sea temperature and PAR doses for east (8 May) and west (26 April) surfaces; cross, bleaching first recorded (14 May).

G. aspera colonies to temperature bleaching in May 1995. First, given that different genotypes of the symbiotic algae on the Caribbean coral *Montastrea annularis* show differential bleaching susceptibility⁶, we tested whether the east and west surfaces of *G. aspera* hosted different algal genotypes. Molecular analysis of the ribosomal RNA gene complex revealed no variation between algae from the two surfaces, either by analysis of restriction-fragment length polymorphisms of the small-subunit-rRNA gene⁶ or by sequence analysis of the ITS1–5.8S–ITS2 region, which resolves finer molecular differences^{7,8}. We conclude that the bleaching pattern observed in *G. aspera* is not due to genetic differences between their algae of the sort previously reported in *M. annularis*⁶.

Second, we investigated whether the east surfaces of *G. aspera* colonies received more solar radiation than the west surfaces at the time of temperature bleaching in May 1995. During solar bleaching (January–March), the dose of PAR was consistently higher on the west than on the east surfaces, but from late April this difference disappeared. The maximal combined sea temperature (31.8–31.9 °C) and solar radiation were recorded on 26 April (for the west surfaces) and 8 May (for the east) — before the first observation (14 May) of temperature bleaching (Fig. 1). So the east surface did not receive more solar radiation than the west at the time of temperature bleaching.

Finally, to establish whether west surfaces are more tolerant to increased temperature, we subjected core samples from the west and east surfaces to different temperatures (elevated, 34 °C; ambient, 27 °C) under identical irradiance for 3 days (Fig. 2). At 27 °C, the algal density and chlorophyll *a* content per algal cell were similar for the east and west cores, and did not vary between the start and end of the experiment. Exposure of the east cores to 34 °C, however, resulted in significantly reduced algal density (Mann–Whitney *U*-test: $P < 0.01$) and caused an increase in the chlorophyll *a* content per algal cell

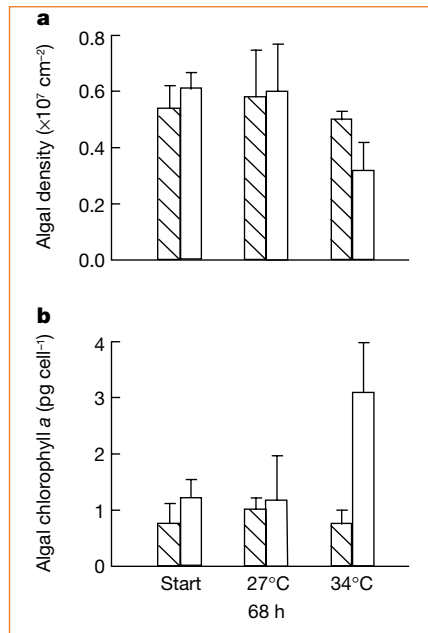


Figure 2 Physiological parameters in east and west cores of *G. aspera* before and after exposure to elevated (34 °C) and ambient (27 °C) temperatures for 68 h at an irradiance of 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Mean values for each parameter are shown; error bars show one standard deviation ($n=5$). **a**, Algal density in west (hatched) and east (clear) cores at the start and end of the experiment at the different temperatures. **b**, Algal chlorophyll *a* concentrations.

($P < 0.008$), whereas west cores were unaffected. Algal chlorophyll *a* was increased in other recently bleached corals^{9,10}.

Our findings indicate that the west surfaces of *G. aspera* colonies could have been protected against temperature bleaching in May 1995 because they were more tolerant than the east surfaces towards the combined stress of temperature and solar radiation. This tolerance probably arose as a result of the increased solar radiation received by the west surfaces during January–March. We conclude that experience-mediated tolerance, as well as algal genotypic differences⁶, contributes to the variation in bleaching susceptibility among reef corals. Many coral colonies are long-lived¹¹, and experience effects may have an important influence on the bleaching responses to global climate change over the coming decades.

B. E. Brown*, **R. P. Dunne***, **M. S. Goodson†**, **A. E. Douglas†**

*Department of Marine Sciences and Coastal Management, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

†Department of Biology, University of York, York YO10 5YW, UK

1. Glynn, P. W. *Coral Reefs* **12**, 1–17 (1993).

2. Brown, B. E. *Coral Reefs* **16**, 129–138 (1997).

3. Brown, B. E., Dunne, R. P. & Chansang, H. *Coral Reefs* **15**, 151–152 (1996).

4. Brown, B. E., Dunne, R. P., Scoffin, T. P. & Le Tissier, M. D. A. *Mar. Ecol. Prog. Ser.* **105**, 219–230 (1994).

5. Brown, B. E. *et al.* *Mar. Ecol. Prog. Ser.* (in the press).

6. Rowan, R., Knowlton, N., Baker, A. & Jara, A. *Nature* **388**, 265–269 (1997).

7. Hillis, D. M. & Dixon, M. T. Q. *Rev. Biol.* **66**, 411–453 (1991).

8. Adachi, M., Sako, Y. & Ishida, Y. *J. Phycol.* **30**, 857–863 (1994).

9. Fitt, W. K., Spero, H. J., Halas, J., White, M. W. & Porter, J. W. *Coral Reefs* **12**, 57–64 (1993).

10. Jones, R. J. *Mar. Ecol. Prog. Ser.* **158**, 51–59 (1997).

11. Hughes, T. P. & Jackson, J. B. C. *Ecol. Monogr.* **55**, 141–166 (1985).

Vision

Myopia and ambient night-time lighting

Myopia is a common affliction (one in four adult Americans is near-sighted¹), and juvenile-onset myopia is believed to be due to a combination of genetic and environmental factors². Results from animal experiments indicate that light cycles may affect the development of myopia^{3,4}, and Quinn *et al.* claim to have extended these to humans⁵. They reported a strong association between childhood myopia and night-time lighting before the age of two: there were five times more children with myopia among those who slept with room lights on than in those who slept in the dark, and an intermediate number among those sleeping with a dim night-light⁵. However, we have been unable to find a link between night-time nursery lighting and the development of myopia in a sample of schoolchildren.

We examined the issue of nursery lighting in a subsample of children from the multicentre Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error (CLEERE) Study. Parents reported their use of night-time lighting and their own refractive status, and the child's refractive error was measured by cycloplegic autorefraction. Our sample consisted of 1,220 children with a median age of 10.2 years: 11.5% of them were African-American, 19.1% Asian, 47.9% Caucasian and 21.6% Hispanic; overall, 18.1% of them were myopic (at least -0.50 dioptres spherical equivalent). The proportion of children with myopia did not differ across nursery-lighting groups ($\chi^2 = 2.62$, $P = 0.271$). Eighty-four of 417 children (20.0%) who slept in darkness were myopic; 128 of 758 children (16.8%) who slept with a night light before age two were myopic, and 10 of 45 children (22.2%) who slept with full room lights on before age two were myopic.

We found an association between the number of myopic parents and nursery lighting before age two ($\chi^2 = 35.02$, $P < 0.001$), as well as an association between ethnicity and room lighting ($\chi^2 = 89.22$, $P < 0.001$). This sample carries a statistical power of 0.99 to be able to detect an odds ratio of 2.00 between nursery lighting and childhood myopia.

Our results do not replicate those of Quinn *et al.*⁵. In fact, the proportion of myopic children in those subjected to a

range of nursery-lighting conditions is remarkably uniform. The association we find between parental myopia and nursery night-time lighting suggests that Quinn *et al.*'s study should have controlled for parental myopia.

Another possible difference is that Quinn *et al.*'s sample is not representative of juvenile myopes. It was drawn from a tertiary referral, paediatric ophthalmology outpatient clinic, and the sample had a median age of eight (young for a sample of myopes) with a very high proportion of myopia (30%). Our sample had fewer myopes and fewer hyperopes, and the children were older. Also, the proportion of parents reporting that their infants slept under full lighting is different in our study: more than 15% of their clinic-based sample had full nursery lighting, whereas only 3.7% of our representative, school-based sample had full room lighting at night.

Our results indicate that myopia is unlikely to develop in children as a result of exposure to night-time lighting as infants.

Karla Zadnik*, **Lisa A. Jones***,
Brett C. Irvin*, **Robert N. Kleinstein†**,
Ruth E. Manny‡, **Julie A. Shin§**,
Donald O. Mutti*, for the CLEERE
Study Group

*College of Optometry, Ohio State University,
Columbus, Ohio 43210-1240, USA

†School of Optometry, University of Alabama,
Birmingham, Alabama 35294-0010, USA

‡College of Optometry, University of Houston,
Houston, Texas 77204-6052, USA

§Southern California College of Optometry,
Fullerton, California 92831, USA

e-mail: zadnik.4@osu.edu

1. Sperduto, R. D., Seigel, D., Roberts, J. & Rowland, M. *Arch. Ophthalmol.* **101**, 405–407 (1983).
2. Mutti, D. O., Zadnik, K. & Adams, A. J. *Invest. Ophthalmol. Vis. Sci.* **37**, 952–957 (1996).
3. Stone, R. A., Lin, T., Desai, D. & Capehart, C. *Vision Res.* **35**, 1195–1202 (1995).
4. Raviola, E. & Wiesel, T. N. *N. Engl. J. Med.* **312**, 1609–1615 (1985).
5. Quinn, G. E., Shin, C. H., Maguire, M. G. & Stone, R. A. *Nature* **399**, 113–114 (1999).

Quinn *et al.* report a strong association between myopia in children and their exposure to night-time lighting during their first two years¹. We have been unable to confirm this surprising result, but we find that myopic parents are more likely to employ night-time lighting aids for their children. Moreover, there is an association between myopia in parents and their children^{2,3}.

We acquired child and parent refraction information as part of a 24-year longitudinal study of visual development in children. These children were research subjects and are not a clinical population. Refractions from 213 children and their parents are included; all children were refracted in the laboratory by non-cycloplegic retinoscopy.

One limitation of Quinn *et al.*'s study is a lack of information about the refractive

status of the parents. Parents in our study were either tested in the laboratory or their spectacle prescriptions were used; if they had never worn glasses and could see clearly at a distance, they were classed as non-myopic.

Subjects (100 females and 113 males) ranged in age from 2 to 24 years, with a mean of 11 years. The data were divided into two groups: myopes, with a spherical equivalent refractive error ranging from -9.0 to -0.5 dioptres (mean, -2.50 dioptres), and non-myopes, with a spherical equivalent refractive error more positive than -0.5 dioptres (range, -0.38 to $+4.38$ dioptres; mean, $+0.87$ dioptres). Answers to questionnaires on nursery lighting conditions at night were collected from parents over the telephone, using the questions of Quinn *et al.* and a few extra ones. One asked parents to rate their confidence in the reliability of their recall of night-time lighting conditions from years earlier: 98% were confident in their responses.

The prevalence of myopia in our sample of children was not associated with ambient light exposure at night during their first two years, or later in life: 20% of those who slept with night lights before age 2 were myopic — the same incidence as in children who slept in the dark. There were no myopes among the small group who slept with full room illumination. This result was not related to either age of onset (mean, 10.5 years) or the severity of myopia.

Families with two myopic parents, however, reported the use of ambient lighting at night significantly more than those with zero or one myopic parent ($\chi^2 = 7.42$, $P < 0.025$). This could be related either to their own poor visual acuity, necessitating lighting to see the child more easily at night, or to the higher socio-economic level of myopic parents, who use more child-monitoring devices. Myopia in children was associated with parental myopia, as reported previously^{2,3}. The proportion of myopic children with two myopic parents was significantly greater than the proportion of myopic children with zero or one myopic parent ($\chi^2 = 4.42$, $P < 0.05$).

Based on these results, we question whether parents need to be concerned about causing myopia in their children by lighting their nurseries at night.

J. Gwiazda, **E. Ong**, **R. Held**, **F. Thorn**
New England College of Optometry, 424 Beacon
Street, Boston, Massachusetts 02115, USA

1. Quinn, G., Shin, C., Maguire, M. & Stone, R. A. *Nature* **399**, 113–114 (1999).
2. Gwiazda, J., Thorn, F., Bauer, J. & Held, R. *Clin. Vision Sci.* **8**, 337–344 (1993).
3. Pacella, R. *et al. Optom. Vision Sci.* **76**, 381–386 (1999).

Quinn *et al.* reply — In not being able to find the strong association reported by us¹ of childhood myopia with night-time ambient lighting before age 2 years, Zadnik *et al.* and

Gwiazda *et al.* ascribe our results to a tendency of myopic parents to illuminate their children's rooms at night. Family studies of myopia typically have difficulty separating environmental from genetic factors, however, as sib-sib correlations for myopia decrease with increasing age difference² and within-family refractive similarities decrease with adjustment for the 'classic' environmental factors of education and close work³. Thus, shared inter-generational behaviour (such as use of night lighting) cannot be excluded *a priori* as contributing to any familial association for myopia.

There are major differences among the studies. Our subjects were younger (mean age, 8 years) and had a considerably higher myopia prevalence of 28% — itself quite high for a United States population of this young age. Accordingly, early-onset myopes, who ultimately tend to become more severely affected, are overrepresented in our tertiary-care population. Thus, it remains to be determined whether the lack of a daily period of darkness during infancy either accelerates myopia onset or provokes the condition in a subset of children who may be predisposed to a more severe form of the condition.

Neither of the subsequent studies considers possible reporting bias. Our findings received widespread publicity, and parents of myopic children might not accurately report or may even under-report a behaviour they fear could have harmed their children. Misclassification errors may also have been introduced into the later results, from non-cycloplegic childhood refractions in one and from self-reported parental refractions⁴ in the other.

Our results¹ and others demonstrating the influence of lighting on ocular development in animals⁵ support the notion that disrupting the daily light-dark illumination cycle may affect eye development in children. Rather than offering reassurance to parents at this time, the disparities in the available clinical reports are better directed to guiding the design of future research into the interactions of light, dark and refractive development.

Richard A. Stone*, **Maureen G. Maguire***,
Graham E. Quinn*†

*Department of Ophthalmology, Scheie Eye
Institute, and †Children's Hospital of Philadelphia,
University of Pennsylvania School of Medicine,
Philadelphia, Pennsylvania 19104-6075, USA
e-mail: stone2@mail.med.upenn.edu

1. Quinn, G. E., Shin, C. H., Maguire, M. G. & Stone, R. A. *Nature* **399**, 113–114 (1999).
2. The Framingham Offspring Eye Study Group *Arch. Ophthalmol.* **114**, 326–332 (1996).
3. Bear, J. C. in *Refractive Anomalies: Research and Clinical Applications* (eds Grosvenor, T. & Flom, M. C.) 57–80 (Butterworth-Heinemann, Boston, 1991).
4. Walline, J. J., Zadnik, K. & Mutti, D. O. *Optom. Vision Sci.* **73**, 376–381 (1996).
5. Stone, R. A., Lin, T., Desai, D. & Capehart, C. *Vision Res.* **35**, 1195–1202 (1995).

Cenozoic motion between East and West Antarctica

Steven C. Cande*, Joann M. Stock†, R. Dietmar Müller‡ & Takemi Ishihara§

* Scripps Institution of Oceanography, La Jolla, California 92093-0215, USA

† California Institute of Technology, Mail Stop 252-21, Pasadena, California 91125, USA

‡ School of Geosciences, Building F05, The University of Sydney, New South Wales 2006, Australia

§ Geological Survey of Japan, 1-1-3 Higashi, Tsukuba, Ibaraki 305-8567, Japan

The West Antarctic rift system is the result of late Mesozoic and Cenozoic extension between East and West Antarctica, and represents one of the largest active continental rift systems on Earth. But the timing and magnitude of the plate motions leading to the development of this rift system remain poorly known, because of a lack of magnetic anomaly and fracture zone constraints on seafloor spreading. Here we report on magnetic data, gravity data and swath bathymetry collected in several areas of the south Tasman Sea and northern Ross Sea. These results enable us to calculate mid-Cenozoic rotation parameters for East and West Antarctica. These rotations show that there was roughly 180 km of separation in the western Ross Sea embayment in Eocene and Oligocene time. This episode of extension provides a tectonic setting for several significant Cenozoic tectonic events in the Ross Sea embayment including the uplift of the Transantarctic Mountains and the deposition of large thicknesses of Oligocene sediments. Inclusion of this East–West Antarctic motion in the plate circuit linking the Australia, Antarctic and Pacific plates removes a puzzling gap between the Lord Howe rise and Campbell plateau found in previous early Tertiary reconstructions of the New Zealand region. Determination of this East–West Antarctic motion also resolves a long standing controversy regarding the contribution of deformation in this region to the global plate circuit linking the Pacific to the rest of the world.

West Antarctica has long occupied an enigmatic position in plate reconstructions. Unlike the East Antarctic craton, which is treated as a coherent block in reconstructions of Gondwanaland, West Antarctica consists of several smaller pieces that have moved relative to each other and to East Antarctica¹. Palaeomagnetic data indicate that there has been at least 300 km of displacement and perhaps 40° to 90° of relative rotation between East Antarctica and West Antarctica since 100 Myr ago^{2,3,4}. The evidence that there has been significant Cenozoic (post 65 Myr) extension is more controversial. Reconstructions have suggested that there has been little (< 50 km) extension in the last 85 Myr (ref. 5). However, if there has been an important episode of Cenozoic extension, it is likely to have occurred in the Ross Sea embayment, separating Marie Byrd Land from East Antarctica. Basins filled with late Cretaceous and younger sediments are found in both the eastern and western parts of the embayment. These basins form part of the West Antarctic rift system which, although largely aseismic, is considered active on the basis of abundant late Cenozoic volcanism and tectonism in the region⁶. Evidence for large amounts of Cenozoic extension include the episodic uplift of the Transantarctic Mountains, starting 50–55 Myr ago^{7,8}, and the development of sedimentary basins with more than 7 km of Cenozoic sediments^{9,10} in the western part of the embayment. Recent drilling near Cape Roberts, Antarctica, has shown that greater amounts of sediments in these basins than previously thought are Oligocene in age¹¹, adding to the evidence for a large Cenozoic extensional event.

Plate tectonic constraints on East–West Antarctica motion since chron 34 (83 Myr; all chron ages are from ref. 12) come from seafloor spreading anomalies and fracture zones generated by Australia–East Antarctica spreading on the southeast Indian ridge (SEIR)^{13,14}, Pacific–West Antarctica spreading on the Pacific–Antarctic ridge^{15,16} and, for the short time between chron 20 (43 Myr) and chron 8 (26 Myr), Australia–Pacific spreading in the Emerald and south Tasman ocean crust (STOC) basins that now straddle and are offset by the Macquarie ridge^{17,18} (Fig. 1). For a plate reconstruction of this region to be considered accurate, plate motions calculated by adding the motion of Australia–East Antarctica, East Antarctica–West Antarctica and Pacific–West Antarctica

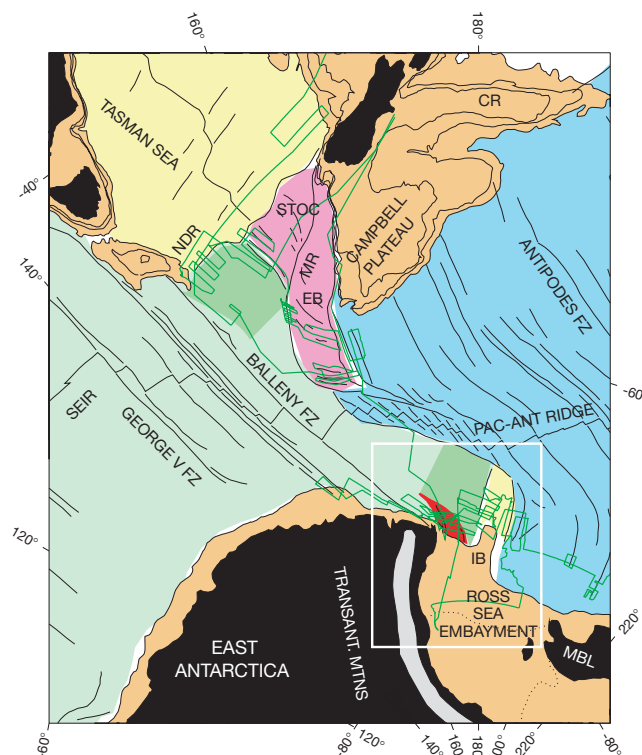


Figure 1 Classification of regions of southwest Pacific ocean crust by the spreading ridge at which they were formed. The two dark green regions east of the Balleny fracture zone were generated by Australia–West Antarctic spreading before chron 8. The red region was generated by East–West Antarctic spreading and the pink region by Australia–Pacific spreading, both between chrons 20 and 8. Light green, blue and yellow indicate regions generated by Australia–East Antarctic, Pacific–West Antarctic and Australia–Lord Howe rise (Tasman Sea) spreading, respectively. Dark green lines show the tracks of the RV Ewing cruise 9513 and RVIB Palmer cruises 9602 and 9702. Abbreviations: CR, Chatham rise; EB, Emerald basin; IB, Iselin bank; MR, Macquarie ridge; NDR, Nella Dan rift; MBL, Marie Byrd Land; SEIR, southeast Indian ridge. The white box shows the location of Fig. 3.

should predict a reasonable plate tectonic history for the Australia–Pacific boundary, and in particular, for the section of that boundary that runs through New Zealand and continues on to the southwest along the Macquarie ridge.

The best evidence for some relative motion between East and West Antarctica in Cenozoic time has been the observation of large misfits in Southwest Pacific reconstructions that do not include motion between East and West Antarctica. The original plate reconstructions for the early Tertiary predicted a large gap between the North and South islands of New Zealand which could only be eliminated by including 500 km of motion between East and West Antarctica¹⁵. Various modifications to the plate tectonic circuit have reduced the amount of this misfit^{19,20}, but the most recent rotation parameters for the SEIR and Pacific–Antarctic ridge still predict a gap of the order of 150 km between the Pacific and Australia plates south of New Zealand in the early Tertiary (Fig. 2a). Sutherland¹⁸ derived a rotation for East–West Antarctica motion by assuming that this gap should be closed. However, his rotation was not well constrained and predicted highly oblique dextral extension in the Ross Sea embayment. Here we present new marine geophysical data from the region north of the Ross Sea and in the South Tasman Sea that allows us to directly calculate East–West Antarctic motion since the early Tertiary period. This motion is tightly constrained by a set of linear seafloor spreading magnetic anomalies, formed in Eocene and Oligocene time, that straddle a fossil spreading centre north of the western Ross Sea embayment.

Balleney fracture zone corridor

We constrained East–West Antarctic motion using the corridor of the southeast Indian ridge east of the Balleny fracture zone. Previous analyses^{14,19} of magnetic and fracture zone data from the SEIR indicated that early Tertiary spreading east of the Balleny fracture zone may have occurred at a slightly faster rate than along the rest of the SEIR (dark green areas in Fig. 1). It was proposed that this discrepancy reflected relative motion between East and West Ant-

arctica, as early Tertiary magnetic anomalies east of the Balleny fracture zone are found north of the Iselin bank, which, in turn, lies to the east of the axes of the Cenozoic basins in the western Ross Sea embayment. More recently, satellite-derived free-air gravity data²¹ revealed the full extent of a graben structure, the Adare trough, north of the western Ross Sea, that had only been partially mapped by conventional methods²² (Fig. 3). This structure was aligned with the western Ross Sea sedimentary basins and seemed likely to be related to East–West Antarctica extension.

Magnetics, gravity, swathmap bathymetry and seismic reflection data were collected in several important areas on the RV *Ewing* cruise 9513 and on the RVIB *Palmer* cruises 9602 and 9702. In addition, several magnetic profiles were collected east of the Balleny fracture zone on transits between Australia and Antarctica by the RV *Hakurei Maru* operated by the Japan National Oil Corporation^{23,24}. Together, these data enable us to quantify the discrepancy in spreading rates and directions across the Balleny fracture zone and to calculate an Euler pole for mid-Cenozoic motion between East and West Antarctica.

Profiles collected on RVIB *Palmer* 9702 across the Adare trough show that a set of magnetic lineations straddles the trough and parallels the graben structure (Fig. 4a). Two profiles (C and D) are characterized by a double peak on the eastern rift flank followed further off axis by a prominent single peak. We identify this sequence as anomalies 11 to 13 and, based on the constant spreading rate model in Fig. 4a, infer that spreading was active between anomalies 18 and 9. The model profile in Fig. 4a indicates that spreading took place at the constant rate of 12.5 mm yr⁻¹, with a slight asymmetry in spreading rate: 5.5 mm yr⁻¹ on the west flank, 7 mm yr⁻¹ on the east flank.

The existence of this anomaly pattern suggests that the Adare trough was one limb of a ridge–ridge–ridge triple junction, and that the Adare trough, not the Balleny fracture zone, was the boundary between East and West Antarctica in Eocene and Oligocene time. To test this idea we analysed the misfits of the SEIR anomalies east of

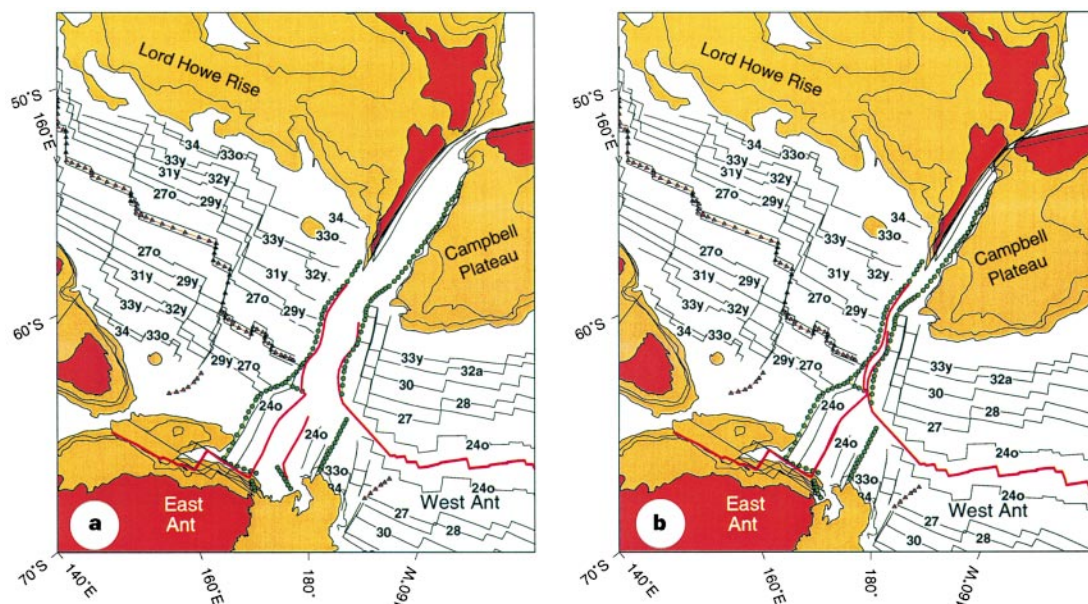


Figure 2 Two reconstructions of the Southwest Pacific for chron 20 constructed by restoring motion along the Pacific–Antarctica–Australia plate circuit. **a**, The plate circuit does not include any East–West Antarctic motion. There is a 150-km gap between the Campbell plateau and Lord Howe rise as well as in the anomaly 20o isochrons (red lines) north of the Ross Sea. **b**, The plate circuit includes an anomaly 20o rotation for East–West Antarctic motion calculated by extrapolating the anomaly 13o rotation as described. The addition of East–West Antarctic motion eliminates the gap in the Pacific–Australia plate

boundary, indicating that the Adare trough is a critical missing plate boundary in the global plate circuit. **a**, **b**, Australia is restored to East Antarctica by a rotation of -24.55° about a pole at 15.07° N, 31.78° E (ref. 25) and the Pacific plate is restored to West Antarctica by a rotation of 35.32° about a pole at 74.78° N, 51.55° W (based in part on unpublished data of S. C. C. and J. M. S). East Antarctica is fixed. Red lines indicate the active plate boundary, brown triangles connected by brown lines indicate fossil spreading centres and green dots show tectonic discontinuities such as triple junction traces and ridge jumps.

the Adare trough that were caused by using rotations based on Australia–East Antarctica anomalies from the SEIR west of the Balleny fracture zone. Revised rotations for the SEIR have recently been calculated for anomalies 20o and 24o (ref. 25) and for anomalies 6o, 8o, 10o and 13o (Cande *et al.*, manuscript in preparation), where the suffix ‘o’ refers to the old end of the anomaly. Unlike previous studies, the revised rotations for anomalies 6o, 8o, 10o and 13o are based on data solely between the Amsterdam–St Paul fracture zone (at 75° E) and the George V fracture zone (at 140° E), thus avoiding potential errors associated with the existence, since at least chron 5, of a distinct Capricorn plate on the north side of the SEIR west of 75° E (ref. 26).

The revised SEIR rotations were used to rotate the anomaly locations from the Australian plate east of the Balleny fracture zone back to the Antarctica plate. These rotations confirm the result of a study¹⁹ that found a misfit in anomalies 13o and 20o east of the Balleny fracture zone (Fig. 5). The rotated position of a small fracture zone that offsets anomaly 13, Fracture Zone A, when compared to its position on the Antarctic plate, indicates that there is a large oblique component to the misfit (Fig. 5). We find that the anomaly 10o and anomaly 13o locations east of the Adare trough are misfit by about 20 km and 75 km, respectively, whereas anomaly 20o locations are misfit by 180 km. Younger anomalies (6o and 8o) show no misfit and nor do the anomaly 10o and 13o identifications between the Balleny fracture zone and the Adare trough.

The most important aspect of Fig. 5 is that the size of the misfit in anomaly 13o is the same as the distance between the anomaly 13o locations across the axis of the Adare trough, confirming our identifications of the Adare trough anomaly sequence. We conclude that the Adare trough anomalies were formed by spreading on the third limb of a ridge–ridge–ridge triple junction between the Australia, East Antarctic and West Antarctic plates that started some time before chron 20 (43 Myr), spread uniformly between chrons 20 and 10 (28 Myr) and had ended by chron 8 (26 Myr). The Adare trough is a fossil rift valley that lay at the axis of the spreading centre when spreading ceased. Because the Adare trough is aligned with the Cenozoic basins of the western Ross Sea, we also conclude that the Adare trough anomalies directly record East–West Antarctic spreading.

The onset of this episode of East–West Antarctic motion can be constrained by analysing the misfit in anomalies older than chron 20. Unfortunately, ambiguity in the identification of anomaly 24 (53 Myr) and older anomalies north of the Iselin bank makes it difficult to determine when this motion started. In Fig. 4b we show magnetic and bathymetric profiles that cross the Nella Dan and Scott rifts, the tectonic boundaries bordering the south Tasman rise and the Iselin bank, respectively, where spreading on the SEIR initiated east of the Balleny fracture zone. Anomalies 21, 24 and 26 are relatively straightforward to identify on the young side of the Nella Dan rift (profiles N1 and N2). A model based on these identifications (‘Model’ in Fig. 4b) indicates that spreading on this section of the SEIR started around chron 29 (64 Myr). On the young side of the Scott rift side, north of the Iselin bank, anomalies 21, 24 and 26 can be identified that have the same spacing as on the Nella Dan rift (profiles S1 and S2); these are our preferred identifications of these anomalies. However, the character of anomalies 24 and 26 on the young side of the Scott rift is not as consistent as on the Nella Dan rift profiles and an alternative interpretation (‘24?’ in Fig. 4b) would move anomaly 24 closer to the Scott rift.

The ambiguity in the anomaly 24 identification affects the calculation of the onset of rifting between East and West Antarctica. With our preferred anomaly 24o identification, the misfit does not increase between anomalies 20o and 24o (Fig. 5), implying that East–West Antarctic motion started around chron 20. With the alternative anomaly 24o identification, there is an additional 25 km

of misfit before chron 20, implying that there was extension in the 10-Myr interval between chrons 24 and 20. However, even with the alternative anomaly identification, the rate of opening between East and West Antarctic before chron 20 (roughly 2.5 mm yr⁻¹) is much slower than in the period between chrons 20 and 8.

East–West Antarctic Motion

We would have liked to determine an Euler pole for East–West Antarctica motion directly from the data on the Adare trough, but it is too short a feature and has no fracture offsets to constrain rotation parameters. However, anomaly 13o locations from the SEIR west and east of the Adare trough constrain Australia–East Antarctic and Australia–West Antarctic motions, respectively, and, when combined with the anomaly 13o locations from the Adare trough, they form a three-plate configuration (Fig. 6) that can be quantitatively solved for East–West Antarctic motion. We assume that no other active plate boundaries have disrupted the anomaly 13 isochrons since they were formed. Deformation of the southeast corner of the Australian plate²⁷ would add to the uncertainty of our calculation, but the lack of a misfit in anomaly 8 and younger anomalies suggests that such deformation has been minor. We used the fitting criteria of ref. 28 and the statistical algorithms of refs 29 and 30 to determine best-fit rotation parameters and 95% confidence limits for all three limbs of the triple junction. The anomaly 13o locations and fracture-zone crossings that constrain this geometry are shown in

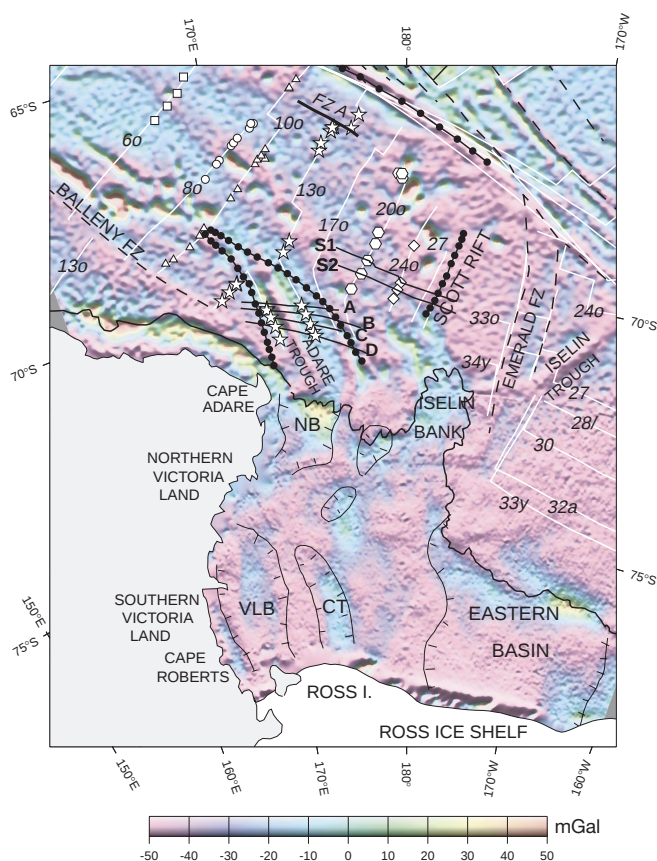


Figure 3 Tectonic features of the western Ross Sea region superimposed on the satellite-derived free-air gravity field. White lines denote isochrons; heavy black line shows 1,000-m contour; black dots indicate tectonic discontinuities. Boundaries of sedimentary basins in Ross Sea are from ref. 10. White filled symbols show control for magnetic isochrons from the Antarctic plate: square, anomaly 6o; circle, 8o; triangle, 10o; star, 13o; hexagon, 20o; diamond, 24o. VLB, Victoria Land Basin; NB, Northern Basin; CT, Central Trough. mGal, milligal. Gravity data are from ref. 21 (north of 72° S) and ref. 37 (south of 72° S).

Table 1 Finite rotations for anomaly 13o three-plate solution

Plate pair	Latitude (+°N)	Longitude (+°E)	Angle (degrees)
Australia & East Antarctica	−13.696	−146.025	20.485
Australia & West Antarctica	−12.600	−147.298	20.848
East Antarctica & West Antarctica	−18.146	−17.847	−0.696

Fig. 6. The fracture-zone crossings were digitized from plots of the satellite-derived free-air gravity field except for Fracture Zone A east of the Balleny fracture zone which is constrained by offsets in shipboard magnetic anomaly data.

Following the statistical method of ref. 30, uncertainties were assigned to the individual anomaly locations and fracture zone crossings. The anomaly locations and fracture zone crossings west of the Balleny fracture zone, which were digitized from Mercator plots, were assigned uncertainties of 7 km and 14 km, respectively. Anomaly locations from east of the Balleny fracture zone, which were digitized using an interactive computer program in which magnetic data points are directly associated with navigation data, were assigned uncertainties of 3 km. The points marking the location of Fracture Zone A were assigned an uncertainty of 9 km. The best-fit rotation parameters for the three plate boundaries are given in Table 1 and the associated covariance matrices constraining the 95% confidence limits are given in Table 2. The value of $\hat{k}$, a parameter that indicates whether the uncertainties are relatively correct, was 0.63. This indicates that the uncertainties assigned to the data points are slightly underestimated. As a check, we note that applying the Australia–West Antarctic rotation to the anomaly 13o

Table 2 Covariance matrices for rotations in Table 1

Plate pair	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
Australia & East Antarctica	3.64	−4.09	2.03	6.16	−3.30	7.78	7
Australia & West Antarctica	5.06	−2.02	7.57	0.835	−2.99	11.4	5
East Antarctica & West Antarctica	2.19	0.0039	5.74	0.0041	0.0083	15.1	5

The covariance matrix is given by the formula

$$\frac{1}{\hat{k}} \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \times 10^{-9}$$

where the values a–f are given in radians squared.

locations from the Australian plate east of the Adare trough eliminates the misfit in these anomalies (Fig. 5).

The 95% confidence region associated with the East–West Antarctica rotation is shown in the inset in Fig. 7. Although this confidence region is large, the rotation provides valuable constraints on the motion between East and West Antarctica. The shape of the confidence region indicates that the geometry of the anomaly 13o locations places a strong constraint on the direction to the Euler pole. A rotation anywhere within the 95% confidence region predicts large east–west extension in the Ross Sea embayment. In Fig. 7 we show the predicted amount of extension, and corresponding 95% confidence ellipses, at several positions in the Ross Sea. There would have been 75 km of extension in a roughly east–west direction for an active plate boundary anywhere in this region with 95% confidence limits on the order of ± 15 km in the northern end of the region and ± 40 km in the southern end of the region. As anomaly 13o is less than 200 kyr younger than the Eocene/

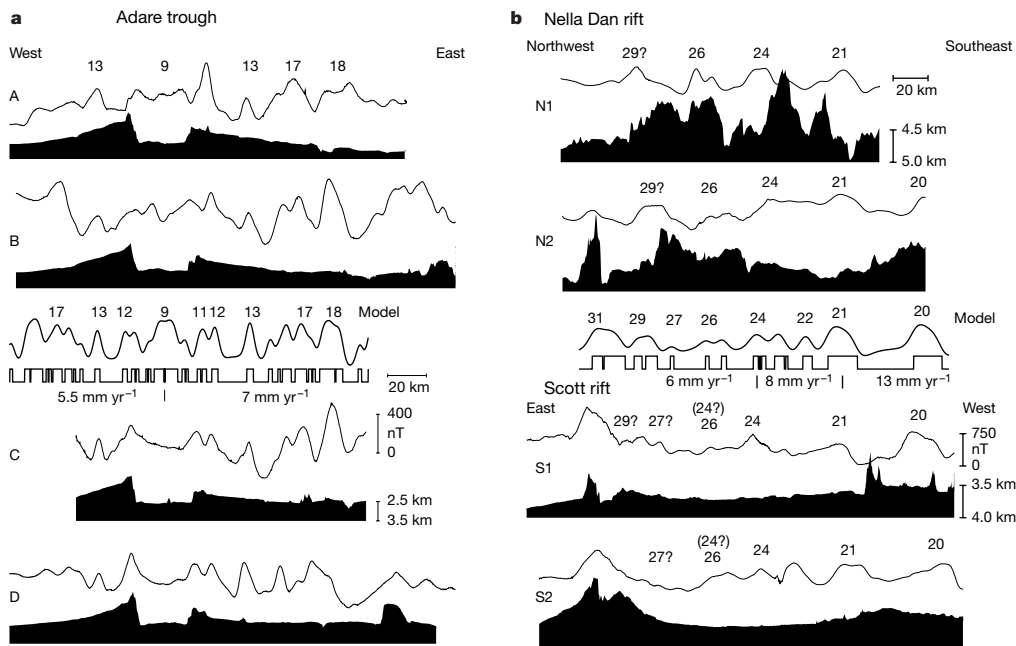


Figure 4 Projected bathymetric and magnetic profiles compared to magnetic models. **a**, Adare trough profiles (A, B, C, D). These data directly constrain the time of East–West Antarctic separation. The model indicates that spreading started around chron 20 and stopped near the end of chron 9. Spreading rates (mm yr^{-1}) are shown beneath the reversal pattern. The model assumes a flat layer between depths of 3.0 and 3.5 km. Other model parameters: present inclination (I_0) = -84° ; present declination (D_0) = -87° ; remanent inclination (I_r) = -70° ; remanent declination (D_r) = 0° ; azimuth = 80° ; magnetization = 7 amps m^{-1} ; gaussian smoothing (σ) = 0.5 km. Location of profiles is

shown in Fig. 3. **b**, Nella Dan rift (N1, N2) and Scott rift (S1, S2) profiles. These data constrain the onset of SEIR spreading east of the Balleny fracture zone. Anomaly 26 is the oldest easily identified anomaly, but spreading clearly started earlier, probably around chron 29. The model assumes a flat layer between depths of 5.0 and 5.5 km. Other model parameters: $I_0 = -76^\circ$; $D_0 = 21^\circ$; $I_r = -70^\circ$; $D_r = 0^\circ$; azimuth = 160° ; magnetization = 7 amps m^{-1} and $\sigma = 1.5 \text{ km}$. Location of Scott rift profiles is shown in Fig. 3, Nella Dan rift profiles in Fig. 6.

Oligocene boundary¹² we take this as a measure of the total Oligocene extension.

The Euler pole seems to be applicable to East–West Antarctic motion back to chron 20o. We estimated a rotation for chron 20o by taking the best-fit East–West Antarctic anomaly 13o pole and gradually increasing the rotation angle in 0.05° increments. At each step we added the new rotation to the Australia–East Antarctic rotation for anomaly 20o (ref. 25) in order to calculate an Australia–West Antarctic anomaly 20o rotation. We found that the anomaly 20o locations from the Australia plate were rotated back to their Antarctic-plate counterparts when the East–West Antarctic rotation angle was 1.70° (Fig. 5). Repeating this exercise for anomaly 24o, we found that no additional increase in the angle was needed to remove the misfits in the preferred anomaly 24o locations, reinforcing our earlier conclusion that the period of spreading started around chron 20o. Although we cannot calculate an uncertainty region for this earlier phase of extension, these observations indicate that there was probably an additional 105 km of separation in Eocene time in the western Ross Sea, to make a total of 180 km of separation in Eocene and Oligocene time.

Implications for the Global Plate Circuit

These results have important implications for the tectonics of Antarctica. The roughly 180 km of Eocene and Oligocene separation provides a tectonic framework for understanding the uplift of the Transantarctic Mountains, the development of the Victoria Land and Northern basins in the western Ross Sea and the abundant rift-related plutonic and intrusive activity that was occurring simultaneously in Northern Victoria Land³¹. The continuing uplift of the Transantarctic mountains in mid-Cenozoic time can be correlated

to motion recorded in the opening of the Adare trough. However, it is less certain whether the onset of Cenozoic uplift at 55–50 Myr, as dated by fission track analyses^{7,8}, is directly related to the axis of extension through the Adare trough region. As we noted, the main episode of extension near the Adare trough seems to have started around chron 20 (43 Myr), 7–12 Myr younger than the onset of uplift. However, as the misfits of anomalies older than anomaly 20 are poorly constrained, it is possible that slow extension between Northern Victoria Land and the Iselin bank started earlier than 43 Myr. If this is the case, then the uplift of the Transantarctic Mountains after 55 Myr might be related to the initiation phase of Adare trough rifting.

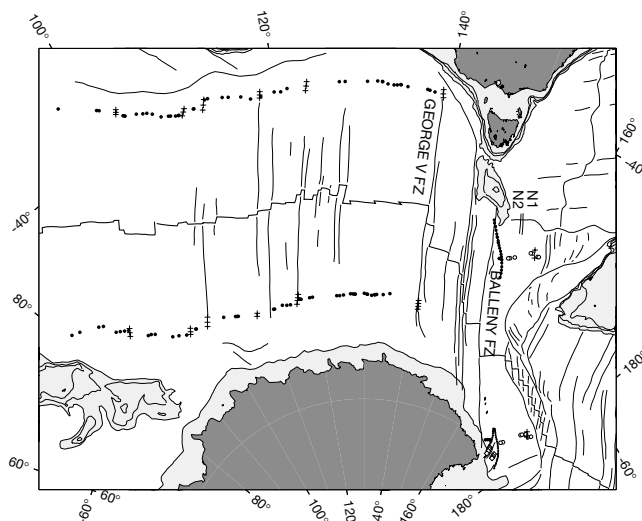


Figure 6 Data constraining the three-plate solution for the Adare triple junction at chron 13o. Anomaly 13o locations for Australia–East Antarctica motion (west of the Adare trough) are shown by filled circles, Australia–West Antarctic motion (east of the Adare trough) by open circles and East–West Antarctic (Adare trough) motion by open diamonds. Fracture zone constraints are shown by + symbols. Profile locations as in Fig. 4. FZ, fracture zone.

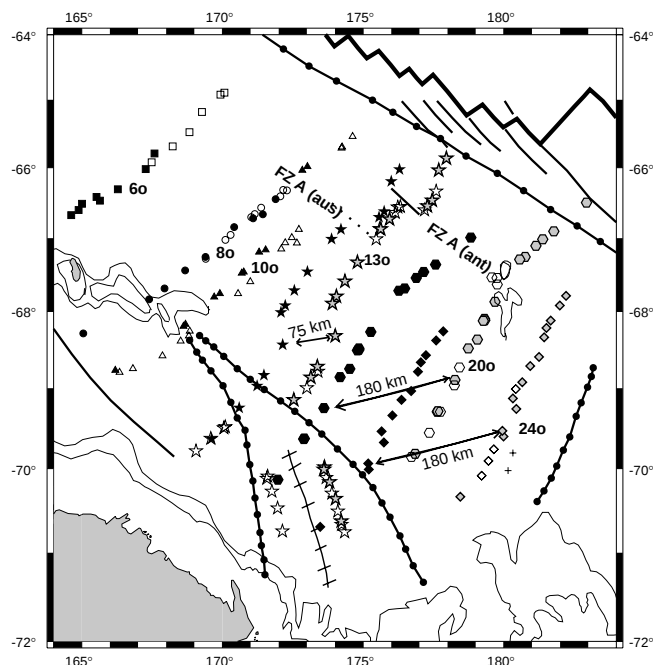


Figure 5 Misfit of anomalies east of the Balleny fracture zone caused by East–West Antarctic motion. Anomalies from the Australian plate (filled symbols) are rotated to the Antarctic plate using Australia–East Antarctic rotations and are compared to the position of Antarctic plate anomalies (open symbols). Australian plate anomalies 10o, 13o and 20o are underrotated by 20, 75 and 180 km, respectively. Shaded anomaly 13o locations have been rotated by the Australia–West Antarctic rotation in Table 1. Shaded anomaly 20o and 24o locations were rotated by Australia–West Antarctic rotations, described in the text. Magnetic anomaly symbols are the same as in Fig. 3 except that the + symbols on the old side of anomaly 24o show the alternative anomaly 24o locations discussed in the text.

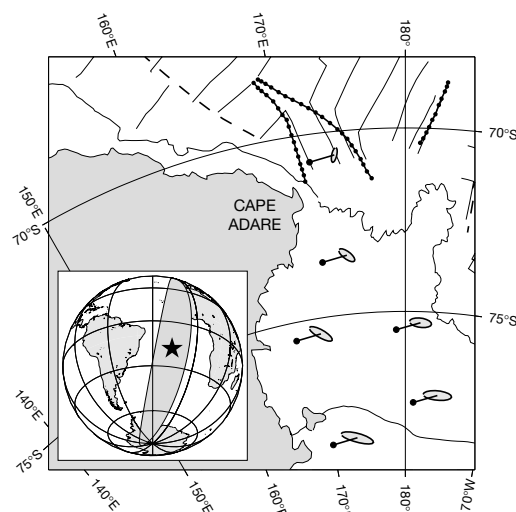


Figure 7 Amount of extension in the western Ross Sea embayment between chrons 13o and 8o (33–26 Myr). Predicted displacements and their 95% confidence ellipses are calculated for points on East Antarctica relative to a fixed West Antarctica. The total amount of mid-Cenozoic extension was probably more than twice this amount, as discussed. Inset, the best-fit East–West Antarctica (Adare trough) anomaly 13o pole (star) and its 95% confidence region.

There is also the possibility of an alternative axis of rifting between East and West Antarctica before chron 24 (53 Myr). The Iselin trough located to the northeast of the Iselin bank (Fig. 3) accommodated up to 50 km of extension between chrons 27 and 24 (Cande *et al.*, manuscript in preparation). If this motion extended to the southwest across the Ross sea to the Transantarctic mountains, there would have been motion between East and West Antarctica associated with it. However, the Iselin trough is nearly at right angles to the Adare trough and the motion along the Transantarctic mountains would probably have been in a NW–SE direction and very oblique. Although there is evidence for a period of oblique motion in South Victoria Land³², this direction of motion seems unlikely to have been capable of generating large extensional uplift.

The termination of Adare trough spreading between chrons 10 and 8 (28 and 26 Myr) corresponds to an important change in southwest Pacific tectonics. In the West Antarctic region, this time marks the onset of widespread volcanism across Marie Byrd Land³³. In the Ross Sea region, it has been proposed³⁴ that 30 Myr corresponds to a fundamental change in the axis of tectonic extension from ENE–WSW to WNW–ESE. South of New Zealand, this time marks the termination of extension between the STOC and Emerald Basins straddling the Pacific–Australia plate boundary¹⁸ and the onset of the current regime of oblique transform motion along the Macquarie ridge.

The recognition of this episode of East–West Antarctic extension has important implications for the tectonics of the entire southwest Pacific plate circuit and, indeed, the global plate circuit. As noted above, a significant tectonic puzzle of the southwest Pacific has been that previous reconstructions for the early Tertiary have predicted a gap between the Lord Howe rise and Campbell plateau. The reconstruction for chron 20 in Fig. 2a, which does not include motion between East and West Antarctica, shows that the gap between the Australia and Pacific plates is comparable to a gap in the anomaly 20a isochrons east of the Balleny fracture zone. Figure 2b is a reconstruction for chron 20 that restores motion between East and West Antarctica using the anomaly 13a best-fit pole with the 1.70° angle we found necessary to eliminate the anomaly 20 misfit east of the Balleny fracture zone. The gap between the Pacific and Australia plates southwest of New Zealand is eliminated, indicating that the Adare trough region is the final critical missing plate boundary in the Southwest Pacific. The inclusion of motion across the Adare trough in the plate circuit will modify the plate motion history associated with the Alpine Fault in New Zealand and also affect global issues such as the motion between hotspots in the Pacific and Indo-Atlantic Oceans^{35,36}. □

Received 30 August 1999; accepted 24 January 2000.

1. Dalziel, I. W. D. & Elliot, D. H. West Antarctica: problem child of Gondwanaland. *Tectonics* **1**, 3–19 (1982).
2. Grindley, G. W. & Olivier, P. J. in *Antarctic Earth Science* (eds Oliver, R. L., James, P. R. & Jago, J. B.) 573–578 (Australian Academy of Science, Canberra, 1983).
3. DiVenere, V. J., Kent, D. V. & Dalziel, I. W. D. Mid-Cretaceous paleomagnetic results from Marie Byrd Land, West Antarctica: A test of post-100 Ma relative motion between East and West Antarctica. *J. Geophys. Res.* **99**, 15115–15139 (1994).
4. Luyendyk, B., Cisowski, S., Smith, C., Richard, S. & Kimbrough, D. Paleomagnetic study of the northern Ford Ranges, western Marie Byrd Land, West Antarctica: Motion between West and East Antarctica. *Tectonics* **15**, 122–141 (1996).
5. Lawver, L. A. & Gahagan, L. M. Constraints on timing of extension in the Ross Sea region. *Terra Antart.* **1**, 545–552 (1994).
6. Behrendt, J. C. *et al.* Geophysical studies of the west Antarctic rift system. *Tectonics* **10**, 1257–1273 (1991).
7. Fitzgerald, P. G., Sandiford, M., Barrett, P. J. & Gleadow, J. W. Asymmetric extension associated with uplift and subsidence in the Transantarctic Mountains and the Ross Embayment. *Earth Planet. Sci. Lett.* **81**, 67–78 (1986).

8. Fitzgerald, P. G. The Transantarctic Mountains in Southern Victoria Land: the application of apatite fission track analysis to a rift-shoulder uplift. *Tectonics* **11**, 634–662 (1992).
9. Cooper, A. K., Davey, F. J. & Behrendt, J. C. in *The Antarctic Continental Margin: Geology and Geophysics of the Western Ross Sea* (eds Cooper, A. K. & Davey, F. J.) Earth Sci. Ser. 5B, 27–76 (Circum-Pacific Research Council, Houston, Texas, 1987).
10. Davey, F. J. & Brancolini, G. in *Geology and Seismic Stratigraphy of the Antarctic Margin* (eds Cooper, A. K., Barker, P. F. & Brancolini, G.) Antarctic Res. Ser. 68, 167–182 (American Geophysical Union, Washington, DC, 1995).
11. Cape Roberts Science Team. Studies from the Cape Roberts Project, Ross Sea, Antarctica - Initial Report on CRP-2/2A. *Terra Antart.* **6**, 1–173 (1999).
12. Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic timescale for the late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).
13. Weisel, J. K. & Hayes, D. E. in *Antarctic Oceanology II: The Australian-New Zealand Sector* (ed. Hayes, D. E.) Antarctic Res. Ser. 19, 165–196 (American Geophysical Union, Washington, DC, 1972).
14. Royer, J. -Y. & Sandwell, D. T. Evolution of the Eastern Indian Ocean since the Late Cretaceous: constraints from GEOSAT altimetry. *J. Geophys. Res.* **94**, 13755–13782 (1989).
15. Molnar, P., Atwater, T., Mamerickx, J. & Smith, S. M. Magnetic anomalies, bathymetry, and the tectonic evolution of the South Pacific since the late Cretaceous. *Geophys. J. R. Astron. Soc.* **40**, 383–420 (1975).
16. Cande, S. C., Raymond, C. A., Stock, J. M. & Haxby, W. F. Geophysics of the Pitman Fracture Zone and Pacific–Antarctic plate motions during the Cenozoic. *Science* **270**, 947–953 (1995).
17. Weisel, J. K., Hayes, D. E. & Herron, E. M. Plate tectonics synthesis: The displacements between Australia, New Zealand, and Antarctica since the Late Cretaceous. *Mar. Geol.* **25**, 231–277 (1977).
18. Sutherland, R. The Australia-Pacific boundary and Cenozoic plate motions in the SW Pacific - some constraints from Geosat data. *Tectonics* **14**, 819–831 (1995).
19. Stock, J. M. & Molnar, P. Uncertainties in the relative positions of the Australia, Antarctica, Lord Howe, and Pacific plates since the Late Cretaceous. *J. Geophys. Res.* **87**, 4697–4714 (1982).
20. Stock, J. M. & Molnar, P. Revised early Tertiary history of plate motion in the Southwest Pacific. *Nature* **325**, 495–499 (1987).
21. Sandwell, D. T. & Smith, W. Marine gravity-anomaly from Geosat and ERS-1 satellite altimetry. *J. Geophys. Res.* **102**, 10039–10054 (1997).
22. Houtz, R. & Davey, F. J. Seismic profiler and sonobuoy measurements in Ross Sea, Antarctica. *J. Geophys. Res.* **78**, 3448–3468 (1973).
23. Sato, S., Asakura, N., Saki, T., Oikawa, N. & Kaneda, Y. Preliminary results of geological and geophysical survey results in the Ross Sea and in the Dumont D'Urville Sea, off Antarctica. *Mem. Natl Inst. Polar Res. (Special Issue)* **33**, 66–92 (1984).
24. Tanahashi, M., Ishihara, T., Yuasa, M., Murakami, F. & Nishimura, A. Preliminary report of the TH95 geological and geophysical survey results in the Ross Sea and Dumont D'Urville Sea. *Proc. NIPR Symp. Antarct. Geosci.* **10**, 36–58 (1997).
25. Tikku, A. A. *Late Cretaceous to Early Tertiary Tectonics of the Southeast Indian Ocean*. PhD thesis, 162, Univ. California, San Diego (1999).
26. Royer, J. -Y. & Gordon, R. G. The motion and boundary between the Capricorn and Australian plates. *Science* **277**, 1268–1274 (1997).
27. DeMets, C., Gordon, R. G. & Argus, D. F. Intraplate deformation and the closure of the Australia–Antarctica–Africa plate circuit. *J. Geophys. Res.* **93**, 11877–11897 (1988).
28. Hellingner, S. J. The uncertainties of finite rotations in plate tectonics. *J. Geophys. Res.* **86**, 9312–9318 (1981).
29. Chang, T. On the statistical properties of estimated rotations. *J. Geophys. Res.* **87**, 6319–6329 (1987).
30. Royer, J. -Y. & Chang, T. C. Evidence for relative motions between the Indian and Australian plates during the last 20 My from plate tectonic reconstructions - implications for the deformation of the Indo-Australian plate. *J. Geophys. Res.* **96**, 11779–11802 (1991).
31. Tonarini, S., Rocchi, S., Armienti, P. & Innocenti, F. in *The Antarctic Region: Geological Evolution and Processes, Proceedings of the VII International Symposium on Antarctic Earth Sciences* (ed. Ricci, C. A.) 511–521 (Terra Antarctica Publication, Siena, 1997).
32. Wilson, T. J. Cenozoic transtension along the Transantarctic Mountains - West Antarctic rift boundary, southern Victoria Land, Antarctica. *Tectonics* **14**, 531–545 (1995).
33. LeMasurier, W. E. in *Volcanoes of the Antarctic Plate and Southern Oceans* (eds LeMasurier, W. E. & Thomson, J. W.) Antarctic Res. Ser. 48, 1–19 (American Geophysical Union, Washington, DC, 1990).
34. Salvini, F., Brancolini, G., Busetti, F. & Mazzarini, F. Cenozoic geodynamics of the Ross Sea region, Antarctica: Crustal extension, intraplate strike-slip faulting, and tectonic inheritance. *J. Geophys. Res.* **102**, 24669–24696 (1997).
35. Molnar, P. & Stock, J. M. Relative motions of hotspots in the Pacific, Atlantic and Indian Oceans since late Cretaceous time. *Nature* **327**, 587–591 (1987).
36. Acton, G. D. & Gordon, R. G. Paleomagnetic tests of Pacific plate reconstructions and for motion between hotspots. *Science* **263**, 1246–1254 (1994).
37. McAdoo, D. & Laxon, S. Antarctic Tectonics: Constraints from an ERS-1 satellite marine gravity field. *Science* **276**, 556–560 (1997).

Acknowledgements

We thank the captains, officers, crews and scientific staff of the RV *Ewing* and RVIB *Palmer* for their dedicated efforts. We also thank B. Luyendyk and R. Sutherland for helpful comments. The project was funded by grants from the National Science Foundation (S. C. C. and J. M. S.).

Correspondence and requests for materials should be addressed to S.C.C. (e-mail: scande@ucsd.edu).

Autoinhibition and activation mechanisms of the Wiskott–Aldrich syndrome protein

Annette S. Kim, Lazaros T. Kakalis, Norzehan Abdul-Manan, Grace A. Liu & Michael K. Rosen

Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

The Rho-family GTPase, Cdc42, can regulate the actin cytoskeleton through activation of Wiskott–Aldrich syndrome protein (WASP) family members. Activation relieves an autoinhibitory contact between the GTPase-binding domain and the carboxy-terminal region of WASP proteins. Here we report the autoinhibited structure of the GTPase-binding domain of WASP, which can be induced by the C-terminal region or by organic co-solvents. In the autoinhibited complex, intramolecular interactions with the GTPase-binding domain occlude residues of the C terminus that regulate the Arp2/3 actin-nucleating complex. Binding of Cdc42 to the GTPase-binding domain causes a dramatic conformational change, resulting in disruption of the hydrophobic core and release of the C terminus, enabling its interaction with the actin regulatory machinery. These data show that ‘intrinsically unstructured’ peptides such as the GTPase-binding domain of WASP can be induced into distinct structural and functional states depending on context.

Proteins in the Ras superfamily of small GTPases regulate a vast array of cellular processes, ranging from growth and differentiation to vesicle trafficking. Members of the family cycle between active GTP-bound and inactive GDP-bound states, and thereby act as molecular switches in signal transduction pathways linking cell-surface receptors to intracellular machinery¹. The Rho subfamily of Ras-like GTPases functions in signalling pathways that control cytoskeletal structure, gene expression, cell-cycle progression and tumour metastasis². The three principal members of this group, Cdc42, Rac and Rho, are each capable of inducing distinct actin-based structures in cells, filopodia, lamellipodia and stress fibres, respectively². Rho acts predominantly through the bundling of pre-existing actin structures, whereas Cdc42 and Rac stimulate *de novo*

actin polymerization, a process that can be mediated by members of the Wiskott–Aldrich syndrome protein (WASP) family³. Mutations in the archetypal member of this group, WASP, lead to the Wiskott–Aldrich syndrome, a paediatric disorder characterized by actin cytoskeletal defects in haematopoietic cells, leading clinically to thrombocytopenia, eczema and immunodeficiency⁴. The WASP proteins signal to the cytoskeleton through the Arp2/3 complex, an actin-nucleating assembly that regulates the structure and dynamics of actin filament networks at the leading edge of the cell³. WASP family members, including WASP, the widely expressed neuronal WASP (N-WASP), several Scar/WAVE proteins and Be1p/Las17p, are characterized by a central segment that is rich in proline followed by a C-terminal VCA region (for verprolin

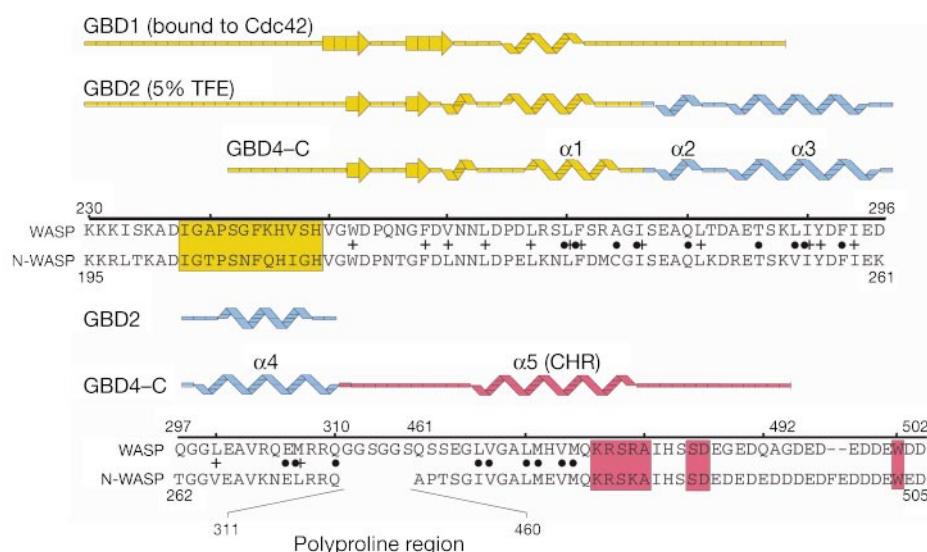


Figure 1 Amino-acid sequences of human WASP and N-WASP with consensus secondary structure of the GBD constructs of WASP shown above. In the sequence, the CRIB motif is boxed in yellow and the conserved Arp2/3 binding sequences are boxed in red. Secondary structures are coloured according to the functional layers described in the text: yellow, layer 1 (residues 230–276); blue, layer 2 (277–310); and red, layer 3 (461–492). Although in our initial report of the Cdc42–WASP complex⁵ we described GBD1 as

forming a β -hairpin with short β -strands 252–253 and 257–258, recent analyses have revealed two additional backbone hydrogen bonds between Gly 251 and Asp 259 and sheet-like NOEs between Val 250 and Val 260, consistent with the longer demarcation. Residues involved in the layer 1/ layer 2 interface (+) and in the layers 1 and 2/ layer 3 (filled circles) are highlighted.

homology region (VHR), cofilin homology region (CHR), and acidic region (AR)). The VCA region binds the Arp2/3 complex and actin, and together these interactions lead to enhancement of actin nucleation by the Arp2/3 complex and rapid formation of new actin filaments.

In contrast to their common carboxy termini, the amino termini of WASP proteins are distinct, enabling family members to activate Arp2/3 in response to different upstream signals. The N termini of WASP and N-WASP contain a 16-residue sequence, termed a CRIB (for Cdc42/Rac interactive binding) motif, which is found in a number of otherwise unrelated effectors of Cdc42 and Rac⁵. Both WASP and N-WASP bind activated Cdc42 through a GTPase-binding domain (GBD) that consists of the CRIB motif and surrounding sequences. We have reported the structure of a Cdc42–WASP GBD complex which reveals that the conserved CRIB residues form a linear epitope at the N terminus that anneals to the β 2 strand of the GTPase, whereas C-terminal residues fold into a β -hairpin and α -helix that pack against Switch I and II (ref. 6). In N-WASP, the GBD is also capable of binding the C-terminal VCA region⁷, resulting in autoinhibition of the Arp2/3-stimulatory activity of the protein^{7,8}. Binding of Cdc42 is believed to cause a structural transition in N-WASP that results in dissociation of its intramolecular contacts and enhanced Arp2/3-mediated actin polymerization^{7,8}. Thus, by activating N-WASP, Cdc42 controls the activity of the Arp2/3 complex and subsequent actin dynamics necessary for formation of filopodia.

Here we report that, like N-WASP, the WASP GBD can bind its own VCA domain in a manner that is competitive with Cdc42. Although the GBD is largely unstructured in the free state, binding of the VCA region or addition of small amounts of organic cosolvent induces formation of a compact, folded domain structure in residues immediately C-terminal to the CRIB motif. In this autoinhibited state, the CHR forms an amphipathic helix that packs against a hydrophobic surface of the GBD, explaining its decreased affinity for actin and the Arp2/3 complex. Notably, this fold of the GBD is sterically incompatible with the activated conformation seen in the Cdc42 complex. Comparison of the conformations and biochemical properties of these two GBD states reveals the structural and thermodynamic basis of activation of WASP-family proteins by Cdc42.

WASP GBD binding to the VCA region

To determine whether WASP and N-WASP can be regulated similarly, we examined the binding of a series of WASP VCA-derived fragments to several GBD constructs. The VCA constructs were prepared as glutathione *S*-transferase (GST) fusion proteins, spanning residues 420–502 (VCA), 420–499 (VC Δ A), 433–499 (Δ VC Δ A), 446–499 (C Δ A), 461–492 (C), 420–492 (VC) and 420–482 (V Δ C). Binding partners for these sequences included GBD1 (residues 230–288, the minimal high-affinity Cdc42-binding domain^{6,9}), GBD2 (residues 230–310), GBD3 (residues 225–310) and GBD4 (residues 242–310, lacking a portion of the CRIB motif) (Fig. 1).

To identify the minimal GBD-binding region of the VCA domain, we carried out an affinity titration assay that was based on the saturable change in tryptophan fluorescence of GBD3 upon the addition of thrombin-cleaved VCA constructs (Fig. 2a). All CHR-containing constructs bound GBD3 with similar affinities, whereas the V Δ C peptide that lacks the latter portion of the CHR proved unable to bind GBD3. These results indicate that the CHR of the VCA domain is both sufficient and necessary for interaction with the WASP GBD.

We then sought to identify the minimal VCA-binding region of the GBD. GBD2, GBD3 and GBD4 all bound immobilized CHR-containing constructs equally well, indicating that the basic region at the N terminus of GBD3 is not required for formation of the complex, in contrast to previous hypotheses⁷ (Fig. 2a; Fig. 2b, lanes

2 and 3). However, C-terminal GBD residues are essential for binding, as GBD1 failed to interact significantly in the GST-affinity assays (Fig. 2b, lanes 4 and 5). These results indicate that formation of the intramolecular complex requires sequences of the GBD that are C-terminal to the CRIB motif, but not the entire CRIB motif.

Binding of the N-WASP GBD to either Cdc42 or to VCA peptides is mutually exclusive⁷. WASP behaves in a similar manner: incubation of the immobilized GST–C–GBD3 complex with Cdc42–GMPPNP (a nonhydrolysable GTP analogue) effects release of the GBD into solution (Fig. 2b, lanes 7 and 8). Cdc42–GDP, which binds WASP only weakly, does not dissociate the GBD from the VCA peptides (Fig. 2b, lane 6). Thus, for both WASP and N-WASP, interactions between the GBD and VCA regions are incompatible with Cdc42 binding, and GTPase-mediated disruption of an autoinhibitory complex probably has an important role in signalling to the cytoskeleton. To elucidate the mechanism of WASP-family activation by Cdc42, we examined the conformational states of both the unbound and VCA-bound GBD.

Induction of a folded domain in the WASP GBD

We and others have reported that isolated GBD peptides of WASP do not assume a single, discrete structure under physiologic

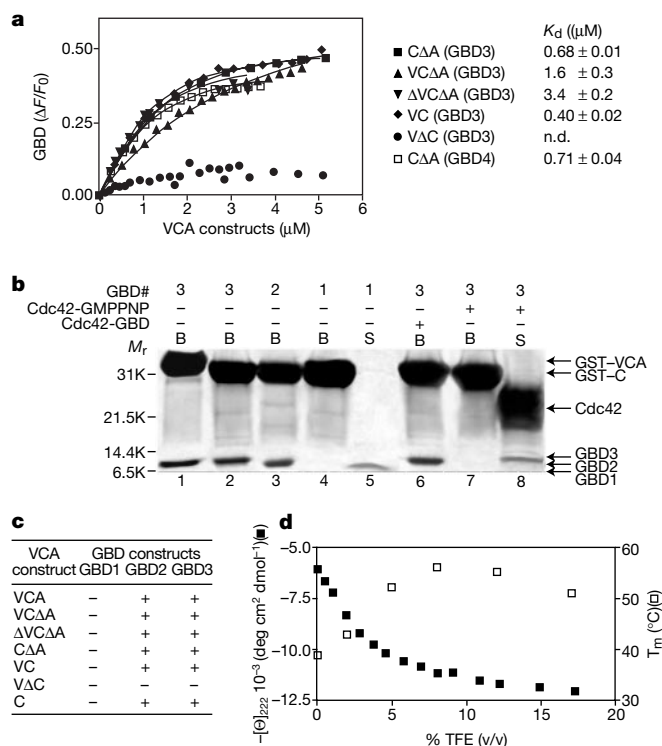


Figure 2 Biochemical characterization of the WASP GBD. **a**, Normalized change in tryptophan fluorescence of GBD3 and GBD4 on addition of VCA peptides. Dissociation constant (K_d) was calculated as described⁹. Tryptophan fluorescence of the GBD was quenched in all cases by ~50% (except in the case of V Δ C) upon binding to thrombin-cleaved VCA constructs that lack tryptophan. n.d., not determined. **b**, Binding of GBD constructs to VCA peptides, and competition by Cdc42. Immobilized GST–VCA (lane 1) or GST–C (lanes 2–8) were incubated with GBD constructs, followed by Cdc42–GDP or Cdc42–GMPPNP as indicated. Proteins retained on the beads ('B' lanes) or released in the supernatant ('S' lanes) were separated by SDS–PAGE and visualized by Coomassie blue staining. Because GBD constructs retain Coomassie blue poorly (A.S.K. and Z.A.M., unpublished data), the relative gel band intensities compared with the GST fusion constructs do not accurately reflect their binding stoichiometry. **c**, Summary of binding data. **d**, Effects of TFE addition on GBD2. Far-ultraviolet CD molar ellipticity recorded at 25 °C at 222 nm (filled squares) and melting temperature (open squares) of 10 μ M GBD2 at different concentrations (v/v) of TFE.

conditions *in vitro*^{6,9}. As expected¹⁰, small amounts of alcoholic solvents (for example, 2,2,2-trifluoroethanol (TFE), methanol, isopropanol) cause a marked increase in the helical content of the peptide as indicated by circular dichroism (CD) spectra, which reaches a plateau at ~10% co-solvent (Fig. 2d). However, this increased helicity correlates with a significant increase in denaturation temperature of the peptide, which rises from 39 °C in the absence of organic solvents to a maximum value of 54 °C at 8–12% TFE (Fig. 2d), a finding more consistent with induction of a folded core than with induction of random helical coils. We reasoned that this fold might mimic that of the GBD–VCA complex.

The NMR spectra of these states supported this hypothesis. Multidimensional spectra of aqueous GBD2 show poor dispersion of amide and aliphatic proton chemical shifts, with large variations in peak shape and intensity (see Supplementary Information). Only ~75% of the expected backbone amide peaks are visible in ¹H/¹⁵N-HSQC (heteronuclear single quantum coherence) spectra, and all amide protons exchange rapidly with solvent. However, the peptide chain is monomeric and relatively compact based on analytical ultracentrifugation and gel-filtration analyses, respectively (not shown). Together, these data suggest that GBD2 samples multiple, loosely packed conformations in aqueous solution.

In contrast, ¹H/¹⁵N-HSQC spectra of ¹⁵N-GBD2 in the presence of 5% TFE and of ¹⁵N-GBD3 in the presence of saturating unlabelled VCA (see Supplementary Information) reveal a marked increase in amide chemical shift dispersion, a decrease in most line widths, development of relative uniformity in peak shapes, and appearance of virtually all expected backbone amide resonances. Under both conditions, a number of backbone amide protons exchange slowly with solvent (14 with TFE, 31 with VCA peptide), based on their observation in ¹H/¹⁵N-HSQC spectra of lyophilized samples freshly dissolved in D₂O + 10% ²H₃-TFE or in D₂O. Significantly, amide and aliphatic resonances are highly similar in the two systems, including ¹H/¹⁵N-HSQC cross-peak patterns and a distinctive upfield methyl resonance at –0.1 p.p.m., which suggests packing of this group against the face of an aromatic ring. This global similarity was preserved in spectra of the minimal GBD4–C complex, with the absence of only several intense, poorly dispersed signals representing disordered residues. These features of the GBD/TFE and GBD/VCA NMR spectra strongly suggest that TFE and the VCA peptide induce a common discrete GBD fold with a well packed hydrophobic core. Notably, GBD spectra recorded in the presence of Cdc42 reveal a very different cross-peak pattern, and the absence of the high field methyl resonance, indicating that the fold of the polypeptide is distinct in the two states (see Supplementary Information).

Autoinhibited structure of the WASP GBD

To understand the nature of the autoinhibited WASP conformation and its relationship to the complex with Cdc42, we determined the NMR solution structures of a GBD–VCA complex and the GBD alone in the presence of 5% TFE. In the former case, we covalently joined the GBD4 and C constructs into a single chain through a flexible (Gly-Gly-Ser)₂ sequence (GBD4–C). This linked construct was monomeric by analytical ultracentrifugation and gel filtration (not shown). Cross peaks in ¹H/¹⁵N-HSQC spectra of GBD4–C nearly superimpose with a subset of peaks in spectra of the larger GBD3–VCA complex and of a wild-type construct spanning residues 230–502 (not shown), indicating that the minimal construct is representative of the intramolecular complex in full-length WASP. Table 1 gives statistics describing the final set of 20 converged conformers for GBD2 in 5% TFE and for GBD4–C. Figure 3 shows the best-fit superpositions of the conformers. Residues in the well-defined core of the TFE structure, 250–310, show 0.83 ± 0.20 Å root-mean-square deviation (r.m.s.d) from the average coordinates for backbone atoms, and 1.33 ± 0.16 Å r.m.s.d. for all heavy atoms, whereas the corresponding 250–310 and 463–478 regions in the

complex display 0.34 ± 0.08 Å r.m.s.d. for backbone atoms and 0.82 ± 0.10 Å r.m.s.d. for all heavy atoms.

In both structures, the N-terminal residues, which encompass the conserved CRIB motif, show no long-range nuclear Overhauser effects (NOEs) until residue 250, poor chemical shift dispersion, and negative or weakly positive backbone amide heteronuclear ¹⁵N{¹H} NOEs (not shown), indicating high levels of mobility in solution. In contrast, C-terminal residues 250–310 form a globular domain consisting of a short *i* + 4 β-hairpin (252–253 and 257–258) and four α-helices (roughly 267–273, α1; 277–281, α2; 284–295, α3; and 298–310, α4; see Figs 1, 4a, c). In these four helices, the φ/ψ angles for the two structures are nearly identical, with minor exceptions at the initial and final residues of α1 and α4. The intramolecular complex also contains a fifth helix, α5, which encompasses residues in the first half of the designated CHR (466–479). Residues after 479 also appear to be mobile in solution on the basis of the criteria listed above. Thus, although residues 483–492 are clearly needed for high-affinity binding to the GBD (Fig. 2a–c), their contribution to binding energy must occur through mechanisms that do not rely on ordering of the peptide backbone.

The β-hairpin, demarcated by two backbone hydrogen bonds between Asp 253 and Gly 257, forms a ~90° angle with the α1 helix, creating an L-shaped amphipathic surface (Fig. 4c). A short turn at residues 259–262 separates the ends of the hairpin and directs the Val 260 side chain toward the core of the domain. The turn is stabilized by a backbone hydrogen bond between the Asp 259 carbonyl and the Asn 262 amide, and by hydrophobic packing of Val 260 against Leu 281 and Thr 282 (Fig. 4c). This first layer of elements packs against a second layer of three roughly planar helices that form three sides of a rectangle, with α3 aligned with the strands of the β-hairpin and α2 and α4 lying perpendicular to them. A hydrogen bond between the amide nitrogen of Phe 258 and the carbonyl oxygen of Leu 281 orders the β-hairpin against the α2 and α3 helices and provides a C-terminal cap for α2. Otherwise, interactions between these two layers are mostly stabilized by hydrophobic contacts, dominated by interactions between Trp

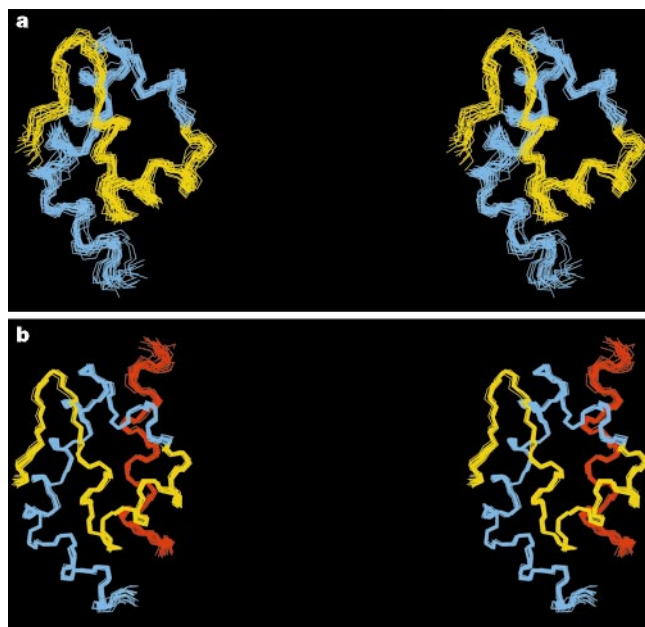


Figure 3 Stereoviews of the best-fit superpositions of the backbone (N, Ca, C) atoms of the 20 final NMR structures of GBD2 in 5% TFE (**a**) and GBD4–C (**b**). Residues 230–249 in **a**, and 241–249, 480–492 and the (Gly-Gly-Ser)₂ linker in **b** are not defined by the NMR data, appear mobile in solution and have been omitted for clarity. Structural layers are coloured as in Fig. 1.

252 and Phe 258 of the β -hairpin; Val 260 and Leu 263 of the β - α 1 turn; Leu 267, Leu 270, Phe 271 and Ile 276 of α 1; Leu 281 of α 2; Ile 290, Tyr 291 and Ile 294 of α 3; and Leu 300, Val 303 and Met 307 of α 4. The close correspondence of the χ_1 rotamers of these residues in the solvent-induced conformation and the intramolecular complex, with the exception only of Tyr 291, shows that the hydrophobic cores of the two structures pack in a nearly identical fashion. This juxtaposition of the first and second layers buries a surface area of $2,180 \pm 70 \text{ \AA}^2$ and places one of the methyl groups of Leu 281 above the aromatic ring of Phe 271, accounting for the large upfield chemical shift of its methyl protons (see Supplementary Information).

The opposite face of the second layer consists of a bed of hydrophobic side chains emanating from helices α 1–4, which are exposed to solvent in the TFE structure. In the intramolecular complex, the α 5 helix of the CHR lies nestled on this bed, forming a third structural layer and completing the core of the domain. Residues Leu 466, Leu 470 and Met 474 of the hydrophobic face of the CHR helix fill three deep pockets on the layer 2 surface, with additional extensive contacts made by Val 467, Met 471, Val 473 and Arg 477 (Fig. 4d). These contacts bury $1,510 \pm 130 \text{ \AA}^2$ of accessible surface between layer 3 and layers 1 and 2. The orientation of the α 5 helix relative to the hydrophobic plane is stabilized by both

electrostatic interactions between the side chains of Arg 477 of the helix and Glu 285 located at the top of the α 2– α 3 turn, and hydrophobic interactions between Leu 270 in the α 1 helix and helices α 3–5. The high conservation of residues at the layer 1/layer 2 and CHR interfaces (Fig. 1) indicates that WASP and N-WASP will share the same autoinhibited fold.

Multiple mechanisms of stabilizing the GBD fold

Preliminary $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shift assignments of aqueous GBD2 indicate that, although the isolated peptide does not exist in a single, discrete, folded conformation, α 1 is fully formed and α 4 is partially formed. In contrast, helical conformations at α 2 and α 3 are not appreciably populated¹¹ (W. Hu and M. K. R., unpublished data). The instability of α 2 and α 3 may be due in part to their enrichment in branched amino acids¹², or to the abundance of acidic residues near their C termini, which could interact unfavourably with their helix dipoles (Fig. 1).

Despite these factors, the autoinhibited fold of the GBD can be induced through two distinct mechanisms, either through the effects of organic co-solvents or by binding to the WASP VCA region. Addition of TFE, a solvent known to induce α -helix formation¹⁰, could stabilize the autoinhibited fold by inducing helical structure in α 2 and α 3. This stabilization of local structure,

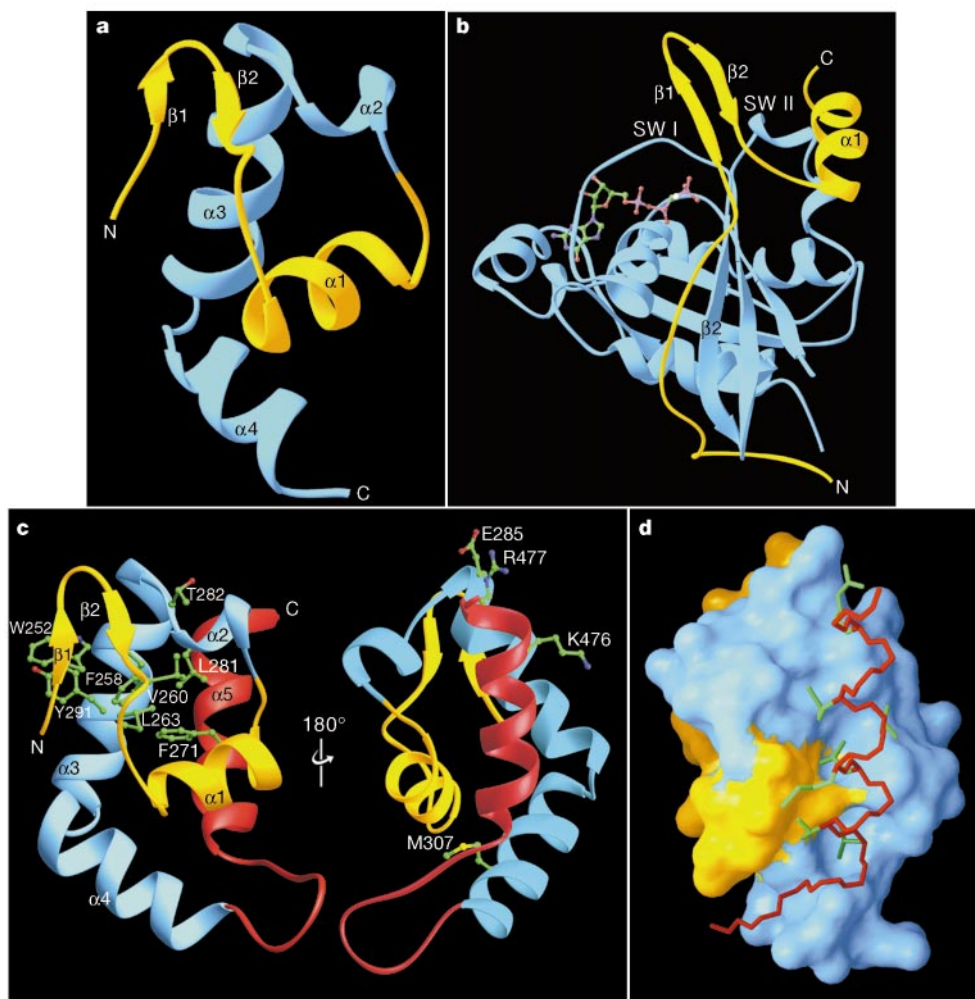


Figure 4 Autoinhibited and activated structures of the WASP GBD. **a**, Ribbons⁴³ depiction of the minimized average coordinates of GBD2 (residues 230–249 not shown) in the presence of 5% TFE. **b**, Structure of GBD1 in complex with Cdc42-GMPPCP (residues 278–288 not shown). Cdc42 is blue, with nucleotide and Mg^{2+} displayed as ball and stick models. SW, switch region. Scale is different from other panels. **c**, Two views of GBD4–C, related by 180° rotation about a vertical axis (residues 241–249 and 480–492 not

shown). Structural layers of the GBD are coloured as in Fig. 1 in all panels. Side chains of residues whose mutation is associated with WASP or XLT are shown. **d**, Solvent-accessible surface of autoinhibited WASP. The cofilin region (layer 3), residues 461–479, was omitted from the surface calculation and is drawn as a stick model. Selected residues in layer 3 that contact layers 1 and 2 are shown. Colouring is as in Fig. 1. Figure generated with GRASP⁴⁴.

and perhaps solvation of the exposed hydrophobic face of layer 2, could in turn facilitate long-range packing and coalescence of the folded core. The VCA region induces structure through a distinct mechanism that may be important in understanding the activation of WASP by Cdc42. As described previously^{6,9}, WASP constructs spanning layer 1 or layers 1 and 2 are not folded in isolation. Similarly, CD data (not shown) indicate that peptides containing layer 3 exist in random coil conformations in solution. By contrast, CD data on GBD4-C (not shown) show a sharp, reversible unfolding transition at 78 °C. Together, these observations indicate that formation of the autoinhibited structure in water may be a cooperative folding process requiring all three layers, involving accretion of secondary structure in helices $\alpha 2$, $\alpha 3$ and $\alpha 5$, and packing of the hydrophobic faces of these elements to form the core of the domain. The related, but mechanistically distinct, folding transitions induced by TFE and VCA peptides illustrate the different environmental factors that can influence the stability and structure of partially folded peptides such as the WASP GBD.

Mechanism of autoinhibition

The autoinhibited fold sequesters the $\alpha 5$ helix through interactions with the GBD. This region of the CHR has been implicated in mediating actin polymerization by the WASP proteins through interaction with the Arp2/3 complex^{7,8,13–15}. Specifically, members of the family are thought to bind the complex in part through a conserved basic motif in the CHR (Lys-Arg-Ser-Basic-Ala in WASP and N-WASP, residues 476–480 and 477–481, respectively)¹⁶. Deletion of Lys 473–Lys 476 in rat N-WASP (Δ cof, corresponding to residues 477–480 of human N-WASP) eliminates the ability of the protein to bind the Arp2/3 complex⁸. Moreover, introduction of this N-WASP mutant into cells inhibits Cdc42-mediated microspike formation⁷, and mutations in the motif are associated clinically with WAS or X-linked thrombocytopenia (XLT) (see below). Together, these data suggest that the basic motif is essential to Arp2/3 binding and biological function of the WASP proteins.

In the autoinhibited GBD structure, residues in the motif lie at the C terminus of the $\alpha 5$ helix (Fig. 4c). Of note, Arg 477 is occluded by layer 2 through a hydrogen bond to Glu 285. Thus, in the intramolecular fold, this residue, and probably residues surrounding it in sequence, would be inaccessible to the binding sites of other partners such as the Arp2/3 complex. In this way, full-length WASP proteins appear to be autoinhibited, activating Arp2/3 only minimally compared to isolated VCA peptides^{8,13}, as a result of sequestration of the basic motif by the intramolecular fold.

Mechanism of activation by Cdc42

Comparison of the autoinhibited and Cdc42-bound structures of the GBD provides insight into the mechanism by which WASP is activated by GTPases. In both the inter- and intramolecular complexes, sequences immediately after the CRIB motif (Val 251–Lys 288) form a similar two-stranded β -sheet and α -helix (Fig. 4a, b). Backbone ϕ and ψ angles throughout this region are virtually identical for all residues in both structures, with the exception of Val 260 and Asn 262. These lie in the α region of ϕ/ψ space in the autoinhibited fold, creating the turn that terminates the $\beta 2$ strand of the GBD, but in the β -region in the Cdc42 complex. Thus, Cdc42 binding eliminates the turn, extracting the Val 260 side chain from Thr 280 and Leu 281 in the core of the domain and placing it against the WASP $\beta 1$ strand. There, Val 260 contacts Val 250, extending the WASP β -strands in the Cdc42 complex to include Val 250–Asp 253 and Gly 257–Val 260 (Figs 1, 4b, 5a, c). The additional hydrogen bonds formed in this restructuring may internally stabilize the WASP peptide, providing some of the driving force for Cdc42 binding. Interactions between the strands may also be strengthened in the complex by sheet-like contacts of $\beta 1$ with Cdc42, in accord with model peptide studies showing cooperative stabilization of three-stranded β -sheets as opposed to β -hairpins¹⁷.

The extension of $\beta 2$ also causes a large change in the relative orientations of the WASP sheet and the $\alpha 1$ helix. In the autoinhibited structure, the helix is roughly perpendicular to the long axis of the sheet, and makes a 135° angle with the short axis (Figs 4c, 5d). Wedged between these two elements are the $\beta/\alpha 1$ loop and the $\alpha 2$ helix. In the Cdc42 complex, this wedge is removed and the $\alpha 1$ helix pivoted such that it is nearly parallel to the long axis of the sheet and packs against one face (Figs 4b, 5c). This conformational compaction, manifested by numerous intra-WASP NOEs, decreases the surface area buried by the layer 1 $\beta/\alpha 1$ sequences from $2,180 \pm 70 \text{ \AA}^2$ in the autoinhibited structure because of packing against layers 2 and 3 (Fig. 5b) to $980 \pm 50 \text{ \AA}^2$ in the complex from packing against switches I and II of Cdc42 (Fig. 5a). The decrease in buried surface in the Cdc42 interface is, however, somewhat compensated by an increase in surface buried within layer 1 between the β -sheet and $\alpha 1$ helix. Notably, virtually all atoms buried in the Cdc42 complex (including those of residues Val 250, Gly 251, Trp 252, Phe 258, Leu 267, Leu 270, Phe 271, Ala 274 and Ile 276) are also buried in the autoinhibited structure and form an integral part of the hydrophobic core (Fig. 5b). Thus, interaction with the GTPase necessitates separation of layer 1 from layer 2, prying open the hydrophobic core of the autoinhibited GBD fold.

This steric incompatibility between the Cdc42-bound and autoinhibited folds contributes to the release of the CHR in two ways. First, when $\alpha 1$ packs against Cdc42, its presence in the CHR binding site is lost (see yellow region in Fig. 4d). Second, separation of layer 1 from layer 2 removes many contacts in the hydrophobic core that stabilize helices $\alpha 2$ and $\alpha 3$. The resulting destabilization of these secondary elements further ensures disruption of the CHR binding site. Thus, as binding of the CHR to one face of layer 2 leads to cooperative folding of the GBD to the autoinhibited conformation, removal of the β -sheet and $\alpha 1$ helix from the opposite face by Cdc42 leads to cooperative unfolding of the domain and release of the

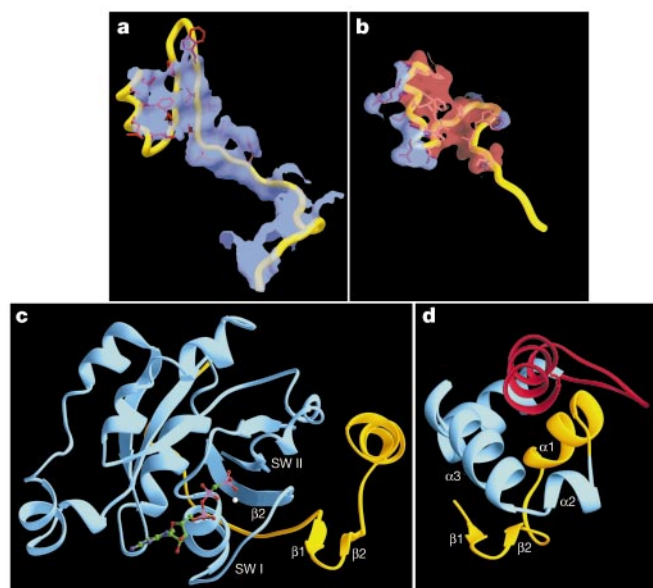


Figure 5 Steric incompatibility of the autoinhibited and activated structures. **a**, Surface of GBD1 that contacts Cdc42 ($d_{c-c} < 4.5 \text{ \AA}$) in the GTPase complex. View is the opposite face of WASP to that shown in Fig. 4b. Backbone is shown as a yellow tube. Side chains from residues 246 to 276 that are part of the interface are red. **b**, Surface (blue) of layer 1 of GBD4-C that contacts layers 2 and 3 in the autoinhibited structure. View is of layer 1 observed from the core of the domain. Surface of atoms that also contact Cdc42 in the activated structure is superimposed in red. Backbone and side chains are represented as in **a**, with a similar orientation of $\alpha 1$. **c, d**, Structures of the GBD1–Cdc42–GMPPCP complex and GBD4–C, respectively, rotated $\sim 90^\circ$ about a horizontal axis relative to Fig. 4b, c. **a, b** generated with GRASP44; **c, d** with Ribbons43.

CHR. The high sequence identity and biochemical similarities between WASP and N-WASP indicate that this mechanism represents the common structural process by which Cdc42 activates these proteins to stimulate the Arp2/3 complex.

Thermodynamics of activation

Our model predicts that Cdc42 should have greater affinity for unfolded WASP constructs than for folded constructs because of the thermodynamic penalty of disrupting the autoinhibited structure of the GBD. To test this prediction, we measured the affinity of a series of GBD constructs for Cdc42 under several conditions using a fluorescence binding assay⁹. In aqueous buffer, GBD1 and GBD2 bind Cdc42 with comparable affinities of 110 ± 7 and 133 ± 9 nM, respectively. As predicted, constructs encompassing both the full CRIB-containing GBD and the full VCA region (FL7, residues 230–502; FL8, residues 230–310 and 420–502 linked by (GGG)₂) show affinities weaker by 83- and 300-fold (11.0 ± 0.7 and 39.7 ± 2.4 μ M), respectively, corresponding to 2.6 and 3.4 kcal mol⁻¹ penalties, probably due to unfolding of the autoinhibited structure upon Cdc42 binding. We obtained similar results for GBD constructs that were folded by the addition of TFE, further supporting our structural model of disruption of the autoinhibited fold upon binding to Cdc42.

Regulation of the GBD–VCA interaction *in vivo*

In addition to Cdc42, other mechanisms may also modulate the intramolecular contacts in WASP and N-WASP *in vivo*, integrating multiple signalling pathways in control of the cytoskeleton. For example, WASP is phosphorylated on tyrosine in response to collagen stimulation of platelets¹⁸, B-cell receptor activation in B cells¹⁹ and IgE receptor activation in mast cells²⁰. A substrate for both Lyn, a Src-family kinase, and Btk, a member of the Tec family, WASP is phosphorylated in a Cdc42-dependent fashion²⁰ at Tyr 291 (mapped for Btk)¹⁹. This tyrosine lies in the $\alpha 3$ helix of the

autoinhibited structure and is largely buried by a π -stacking interaction with Trp 252 in the β -hairpin (Fig. 4c). Disruption of the $\alpha 3$ -hairpin contacts upon Cdc42 binding would greatly increase accessibility of Tyr 291 to kinases, and might explain the potentiating effect of the GTPase on Lyn-mediated phosphorylation. Moreover, phosphorylation of Tyr 291 could destabilize the autoinhibited fold, acting to extend the lifetime of the WASP signal in a GTPase-independent manner. This could occur through either direct destabilization of $\alpha 3$ -hairpin contacts, or by creating a docking site for SH2 domains, whose binding would be incompatible with the helical conformation of $\alpha 3$. Finally, destabilization of the GBD by phosphorylation could affect the rate of WASP turnover, again changing the time course of the actin polymerization signal.

Both SH3-domain-containing proteins, such as Grb2 (M.-F. Carlier, personal communication), and the lipid second messenger PIP₂ (ref. 8) synergize with Cdc42 to enhance stimulation of Arp2/3-mediated actin polymerization by N-WASP. The biochemical mechanisms underlying these effects are unknown, but may involve modulation of the autoinhibited fold. Because the polyproline region of WASP lies between the GBD and VCA sequences, SH3-containing adapter proteins that bind to this region could regulate the equilibrium between the open (activated) and closed (autoinhibited) forms of WASP. Likewise, PIP₂ might act through another type of conformational toggle, affecting the sequestration or structure of the VCA region. The NMR data and methods described above will provide a powerful means of addressing these issues in the future.

WASP in disease

A large database now exists of WASP mutations that cause classical WAS or XLT, a milder form of the disease limited to platelet abnormalities⁴. A number of missense mutations can now be examined in the context of the activated and autoinhibited WASP structures (Fig. 4c). Two of these are in the GBD: M307V (ref. 21) and A236G (ref. 22). Methionine 307 is located in the $\alpha 4$ helix, where it packs against $\alpha 1$. Its mutation appears to destabilize the autoinhibited structure (A.S.K., unpublished data), leading to improper WASP activation or perhaps enhanced proteolytic turnover²³ *in vivo*. On the other hand, Ala 236 is located near the N terminus of the CRIB motif, and in the Cdc42 complex structure it is positioned over the C-terminal helix of the GTPase⁶. Its mutation may directly impact the WASP–Cdc42 interaction, thereby affecting the efficiency of WASP activation by the GTPase.

Two disease-causing mutations within the VCA region, R477K (ref. 24) and K476E (ref. 25), have also been identified (Fig. 4c). Both altered residues lie within the conserved basic motif of the CHR. Although mutation of Arg 477 might be expected to change its interaction with Glu 285 (see above), Lys 476 extends from the solvent-exposed surface of the autoinhibited structure and does not directly make intramolecular contacts. Notably, introduction of each mutation individually into GBD4–C does not change the melting temperature of the construct measured by CD or the ¹H/¹⁵N-HSQC dispersion pattern (not shown), suggesting that neither affects the stability of the autoinhibited fold. Thus, it is likely that the mutations affect interaction of the cofilin helix with another partner. Indeed, residues 476 and 477 are deleted in the Δ cof mutant which fails to bind the Arp2/3 complex⁸, suggesting that their alteration might affect WASP interactions with this assembly. Further biochemical analyses of these various mutant proteins, interpreted in light of the activated and autoinhibited structures, will provide additional insights into the molecular basis of the Wiskott–Aldrich syndrome.

Generalities of WASP in signal transduction

The activation of WASP has many parallels to that of other Ras superfamily effectors, including the kinase Raf, a target of Ras, and the kinase PAK, a Cdc42/Rac target. In each case, binding of the

Table 1 Structural statistics

	WASP GBD2 (5% TFE) (residues 230–310)	WASP GBD4–C (H ₂ O) (residues 242–310, 461–492)
Restraints for structure calculation*		
Total restraints	1,237	2,713
Total NOE restraints	1,107	2,494
Unambiguous	1,006	1,942
Ambiguous	101	552
Dihedrals (ϕ , ψ , χ^1)	43,43,40	69,68,45
Statistics for structure calculation†		
R.m.s.d. from experimental restraints		
Distances (Å)	0.027 $\pm$ 0.001	0.011 $\pm$ 0.002
Angles (°)	0.49 $\pm$ 0.13	0.28 $\pm$ 0.03
Deviations from idealized geometry		
Bonds (Å)	0.0037 $\pm$ 0.0004	0.0015 $\pm$ 0.0002
Angles (°)	0.47 $\pm$ 0.02	0.322 $\pm$ 0.008
Final energies (kcal mol ⁻¹)		
E_{total} , E_{NOE} , E_{cdh}	424 $\pm$ 71, 60 $\pm$ 5, 4 $\pm$ 2	86 $\pm$ 10, 18 $\pm$ 7, 0.9 $\pm$ 0.2
Coordinate precision‡		
R.m.s.d. of backbone atoms	0.83 Å $\pm$ 0.20	0.34 Å $\pm$ 0.08
(N, C α , C β)		
R.m.s.d. of all heavy atoms	1.33 Å $\pm$ 0.16	0.82 Å $\pm$ 0.10
Procheck analysis (residues 250–310) (residues 250–310, 463–479)		
Most favoured	67.4%	94.5%
Additional allowed	25.5%	5.0%
Generously allowed	4.9%	0.3%
Disallowed region	2.2%	0.2%

* The final values for the force constants used for the various terms in the target function used for simulated annealing are as follows: 5 kcal mol⁻¹ Å⁻⁴ for the quartic van der Waals repulsion term (with van der Waals radii set to 0.78 times their value defined in the paramatldhg-pro parameter set (in XPLOR and CNS)); 50 kcal mol⁻¹ Å⁻² for the experimental distance restraints; and 200 kcal mol⁻¹ rad⁻² for the experimental torsion angle restraints.

† None of the final structures have distance restraint violations greater than 0.3 Å or dihedral violations greater than 5°.

‡ Defined as average root-mean-squared difference for the stated residues (see text) between the final 20 structures and the average coordinates.

GTPase to target is mediated by a conserved region (the Ras-binding domain of Raf²⁶, or the CRIB region of the Rho-family targets⁵), but these interactions do not themselves appear to cause the structural changes necessary for effector activation. Such changes are mediated by flanking sequences (the cysteine-rich domain of Raf²⁷, and segments C-terminal to the CRIB motif in WASP and PAK²⁸), whose interactions with the GTPase relieve autoinhibitory contacts to activity-bearing regions of the effectors. In all cases, the non-covalent interactions with the GTPase potentiate homologous (PAK²⁹) or heterologous (WASP^{18–20} and Raf³⁰) phosphorylation events that appear to lead to permanent enhancement of activity. These parallels suggest a common theme of GTPase effector regulation involving the steps of autoinhibition, transient allosteric activation, and covalent modification that extends the time period of activation. The ability of effectors to respond to both transient and permanent activation signals allows them to integrate information from multiple sources, and produce different responses depending on the relative strength, timing, duration and localization of the inputs.

In addition, the requirement for conformational change to affect release of autoinhibition suggests a common theme of protein plasticity in signal transduction. The structures described here identify a single polypeptide chain, the WASP GBD, that can adopt related but distinct folds depending on context (Fig. 4a–c). The structures have similar layer 1 secondary elements which are, however, uniquely packed and oriented when induced by Cdc42, TFE or the VCA region (compare Fig. 5c and d). Adjacent regions, including the CRIB motif and layer 2, are likewise structurally distinct in the different states. It is probable that this extreme plasticity of the GBD is related to its absence of discrete tertiary structure in isolation. Numerous examples have appeared recently in the literature where unfolded, yet fully functional, proteins or protein domains are induced to discrete structure on binding their targets^{31,32}. This coupling of folding and binding may affect ligand specificity, regulation, transport and *in vivo* turnover^{31,32}. Implicit in many functional arguments is the notion that the absence of a single thermodynamic minimum may enable such 'intrinsically unstructured' chains to adopt different minima, characterized by different three-dimensional structures, when bound to different targets or under different conditions^{31,32}. While it remains an open issue whether intrinsically unstructured domains/proteins truly exist in a free, unfolded state in the cell, it is likely that *in vivo* the WASP GBD alternates between the various induced structures discovered here in the context of different binding partners. In either case, the ability of a polypeptide to adopt functionally distinct structures in different environments, which themselves may be controlled by additional cellular inputs, could provide a novel means of regulating information transfer in signaling pathways *in vivo*. The unique biophysical properties of the WASP GBD, coupled with the complex yet dissectable activation process of the protein, make WASP an ideal system for examination of the fundamental physical and biological principles underlying signal transduction in the cell. □

Methods

Sample preparation

Most constructs were amplified from the full-length WASP complementary DNA by PCR and subcloned into either pET11a (GBD1–4), pET16b (FL7) or pGEX-2T (VCA constructs) expression vectors. Proteins were overexpressed in *Escherichia coli* strain BL21(DE3) and purified from bacterial lysate by either conventional chromatography (GBD1–4) or by affinity chromatography on Ni-NTA resin (FL7) or glutathione sepharose (VCA constructs). Isolated VCA proteins were obtained from the GST fusions by cleavage with thrombin and purification by gel-filtration chromatography. GBD4–C and FL8 were constructed by a three-step process in which two separate fragments of the WASP cDNA were amplified by PCR with insertion of a complementary sequence encoding the (Gly-Gly-Ser)₂ linker, followed by PCR amplification of a mixture of those templates. The constructs were subcloned into the pET11a expression vector then overexpressed and purified as above for the GBD proteins.

Uniform ¹⁵N and ¹³C labelling was achieved by growing bacteria in minimal media supplemented with ¹⁵NH₄Cl and/or ¹³C₆-glucose as the sole nitrogen and carbon sources, respectively.

Fluorescence spectroscopy

Fluorescence studies on the VCA constructs were conducted on a Perkin Elmer LS50B Luminescence Spectrometer at 25 °C using a 10-mm rectangular quartz cuvette. The cell was charged with 1 μM of either GBD3 or GBD4 and titrated with purified thrombin-cleaved VCA peptides. The decrease in fluorescence emission at 347 nm subsequent to excitation at 288 nm of the single GBD tryptophan 252 was followed. VCA peptides do not contain tryptophan. Equilibrium dissociation constants as well as the maximum change in fluorescence and the concentration of the VCA peptides were obtained by fitting the solution of the quadratic equation describing a bimolecular association process⁹. Affinity of GBD constructs for Cdc42–(Mant)GMPNP was determined similarly by following the decrease in Mant fluorescence at 466 nm ($\lambda_{\text{excitation}} = 350 \text{ nm}$) on addition of ligand⁹. Measurements were made on a Fluoromax-2 spectrofluorimeter (Spex Inc.) at 25 °C, with Cdc42 concentration between 50 and 1,000 nM. Cdc42 was uniformly loaded with nucleotide as described⁹. All dissociation constant (K_d) values are based on a minimum of three independent titrations.

Binding assays

GST–VCA fusion constructs were immobilized on glutathione S-sepharose beads and incubated with GBD protein (3-fold molar excess) for 2 h at 4 °C. After sedimentation by centrifugation, the beads were washed with buffer and analysed, in parallel with supernatant, by SDS–PAGE. Alternatively, after washing the beads were treated with either GDP- or GMPNP-loaded Cdc42 (10-fold molar excess over GBD). The beads and supernatant were then analysed separately by SDS–PAGE.

Circular dichroism

Circular dichroism spectra were acquired from 10 μM protein solutions using a Jasco 710 spectrometer. Thermal unfolding was monitored by measuring CD at 222 nm as temperature was increased at 1 °C min^{–1} from 10 °C to 90 °C. Melting temperature was determined from the first derivative of the unfolding curve. Identity of the heating and cooling curves indicated reversible unfolding.

NMR spectroscopy

NMR experiments were carried out at 25 °C on Varian Inova 500 or 600 MHz spectrometers equipped with four channels and pulsed field gradients. Samples contained 1–2 mM protein in 50 mM phosphate buffer (90% H₂O/10% D₂O or 100% D₂O) pH 6.8 (uncorrected) and 100 mM NaCl. Backbone ¹HN, ¹⁵N, ¹³Ca, ¹Hα, ¹³Cβ and ¹³CO assignments were obtained for both GBD2 and GBD4–C peptides through the following triple resonance spectra recorded on ¹⁵N/¹³C-labelled samples: HNCO, HNCaCb, CbCa(CO)NNH, HNCaHa, and HbHaCbCa(CO)NNH (ref. 33). Side-chain resonances were assigned using the CCC-TOCSY-NNH, HCC-TOCSY-NNH (ref. 34) and HCCH-TOCSY (ref. 35) experiments. Aromatic protons were assigned through 2D homonuclear ¹H-NOESY and ¹H-TOCSY spectra.

Distance restraints were obtained from 3D and 4D ¹⁵N- and/or ¹³C-edited NOESY spectra, and 2D ¹H-NOESY spectra, recorded with mixing times between 60 and 120 ms. A total of 47/64 (GBD2/GBD4–C) ³J_{HN–Ha} coupling constant restraints were obtained from HNHa experiments³⁶. Stereospecific Hβ assignments and 45/40 (GBD2/GBD4–C) χ₁ restraints (± 20–40°) were obtained from HNHb and HN(CO)Hb experiments and analysis of short-range NOE patterns. Stereospecific assignments of valine and leucine methyl groups were obtained from constant time ¹H/¹³C-HSQC data acquired on 10% ¹³C-labelled samples³⁷. A total of 42/69 (GBD2/GBD4–C) ϕ and 44/68 ψ restraints were included based on analysis of ¹HN, ¹⁵N, ¹³CO, ¹³Ca and ¹³Cb chemical shifts in the program TALOS¹¹. Dihedrals were restrained to the maximum of 30° or twice the standard deviation observed in the TALOS database matches. Steady-state heteronuclear ¹⁵N{¹H} NOEs were measured as described³⁸.

Structure calculations

Initial structures were calculated using a standard torsion angle simulated annealing dynamics protocol implemented in the program CNS version 0.3 (ref. 39). Restraints were classified as strong (1.8–2.7 Å, 2.9 Å for amides), medium (1.8–3.5 Å) and weak (1.8–5.0 Å). An additional 0.5 Å was added to the upper bound for all methyl groups. The structure of GBD2 was refined through iterative, manual assignment of NOEs. That of GBD4–C was refined from an initial global fold in an automated manner using ARIA⁴⁰ in conjunction with X-PLOR (version 3.851)⁴¹. The cut-off value of an ambiguous assignment was reduced progressively from 0.99 for the first iteration to 0.85 in iteration 7. Quality of the structures was evaluated using ProcheckNMR⁴².

Received 8 November 1999; accepted 17 January 2000.

- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125–131 (1990).
- Hall, A. Rho GTPases and the actin cytoskeleton. *Science* **279**, 509–514 (1998).
- Machesky, L. M. & Insall, R. H. Signaling to actin dynamics. *J. Cell Biol.* **146**, 267–272 (1999).
- Derry, J. M. J., Ochs, H. D. & Francke, U. Isolation of a novel gene mutated in the Wiskott–Aldrich syndrome. *Cell* **78**, 635–644 (1994).
- Burbelo, P. D., Drechsel, D. & Hall, A. A conserved binding motif defines numerous candidate target proteins for both Cdc42 and Rac GTPases. *J. Biol. Chem.* **270**, 29071–29074 (1995).
- Abdul-Manan, N. *et al.* Structure of Cdc42 in complex with the GTPase binding domain of the Wiskott–Aldrich syndrome protein. *Nature* **399**, 379–383 (1999).
- Miki, H. *et al.* Induction of filopodium formation by a WASP-related actin-depolymerizing protein N-WASP. *Nature* **391**, 93–96 (1998).
- Rohatgi, R. *et al.* The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221–231 (1999).

9. Rudolph, M. *et al.* The Cdc42/Rac interactive binding region motif of the Wiskott Aldrich syndrome protein (WASP) is necessary but not sufficient for tight binding to Cdc42 and structure formation. *J. Biol. Chem.* **273**, 18067–18076 (1998).
10. Buck, M. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Quart. Rev. Biophys.* **31**, 297–355 (1998).
11. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
12. Rohl, C. A., Chakrabartty, A. & Baldwin, R. L. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **5**, 2623–2637 (1996).
13. Machesky, L. M. *et al.* Scar, a WASP-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proc. Natl Acad. Sci. USA* **96**, 3739–3744 (1999).
14. Winter, D., Lechler, T. & Li, R. Activation of the yeast Arp2/3 complex by Bee1p, a WASP-family protein. *Curr. Biol.* **9**, 501–504 (1999).
15. Yazar, D. *et al.* The Wiskott–Aldrich syndrome protein directs actin-based motility by stimulating actin nucleation with the Arp2/3 complex. *Curr. Biol.* **9**, 555–558 (1999).
16. Bi, E. & Zigmond, S. H. Actin polymerization: where the WASP stings. *Curr. Biol.* **9**, R160–R163 (1999).
17. Sharman, G. J. & Searle, M. S. Cooperative interaction between the three strands of a designed antiparallel β -sheet. *J. Am. Chem. Soc.* **120**, 5291–5300 (1998).
18. Oda, A. *et al.* Collagen induces tyrosine phosphorylation of Wiskott–Aldrich syndrome protein in human platelets. *Blood* **92**, 1852–1858 (1998).
19. Baba, Y. *et al.* Involvement of Wiskott–Aldrich syndrome protein in B-cell cytoplasmic tyrosine kinase pathway. *Blood* **93**, 2003–2012 (1999).
20. Guinamard, R. *et al.* Tyrosine phosphorylation of the Wiskott–Aldrich syndrome protein by Lyn and Btk regulated by CDC42. *FEBS Lett.* **434**, 431–436 (1998).
21. Kwan, S. *et al.* Scanning of the Wiskott–Aldrich syndrome (WAS) gene: identification of 18 novel alternations including a possible mutation hotspot at ARG86 resulting in thrombocytopenia, a mild WAS phenotype. *Hum. Mol. Genet.* **4**, 1995–1998 (1995).
22. Villa, A. *et al.* X-Linked thrombocytopenia and Wiskott–Aldrich syndrome are allelic diseases with mutations in the WASP gene. *Nature Genet.* **9**, 414–417 (1995).
23. Shcherbina, A., Rosen, F. S. & Remold-O'Donnell, E. WASP levels in platelets and lymphocytes of Wiskott–Aldrich syndrome patients correlate with cell dysfunction. *J. Immunol.* **163**, 6314–6320 (1999).
24. Derry, J. M. J. *et al.* WASP gene mutations in Wiskott–Aldrich syndrome and X-linked thrombocytopenia. *Hum. Mol. Genet.* **4**, 1127–1135 (1995).
25. Kolluri, R. *et al.* Identification of WASP mutations in patients with Wiskott–Aldrich syndrome and isolated thrombocytopenia reveals allelic heterogeneity at the WAS locus. *Hum. Mol. Genet.* **4**, 1119–1126 (1995).
26. Block, C. *et al.* Quantitative structure-activity analysis correlation Ras/Raf interaction in vitro to Raf activation *in vivo*. *Nature Struct. Biol.* **3**, 244–251 (1996).
27. Campbell, S. *et al.* Increasing complexity of Ras signaling. *Oncogene* **17**, 1395–1413 (1998).
28. Zhao, Z. S. *et al.* A conserved negative regulatory region in alphaPAK: inhibition of PAK kinases reveals their morphological roles downstream of Cdc42 and Rac1. *Mol. Cell. Biol.* **18**, 2153–2163 (1998).
29. Manser, E. *et al.* A brain serine/threonine protein kinase activated by Cdc42 and Rac1. *Nature* **367**, 40–46 (1994).
30. Stokoe, D. *et al.* Activation of Raf as a result of recruitment to the plasma membrane. *Science* **264**, 1463–1467 (1994).
31. Spolar, R. S. & Record, M. T. Jr Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777–784 (1994).
32. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
33. Yamazaki, T. *et al.* A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N , ^{13}C , ^2H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655–11666 (1994).
34. Logan, T. M. *et al.* A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomol. NMR* **3**, 225–231 (1993).
35. Kay, L. E. *et al.* A gradient-enhanced HCCH-TOCSY experiment for recording side-chain proton and carbon-13 correlations in water samples of proteins. *J. Magn. Reson. Ser. B* **101**, 333–337 (1993).
36. Kuboniwa, H. *et al.* Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR* **4**, 871–878 (1994).
37. Neri, D. *et al.* Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* **28**, 7510–7516 (1989).
38. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
39. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
40. Nilges, M. *et al.* Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J. Mol. Biol.* **269**, 408–422 (1997).
41. Brünger, A. T. *X-PLOR manual*. (Yale Univ. Press, New Haven, 1993).
42. Laskowski, R. A. *et al.* AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
43. Carson, M. J. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).
44. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank M. Lu for analytical ultracentrifugation experiments; W. Hu for communicating results before publication; M. Buck, Y.M. Chook, J. Goldberg, B. Schulman, B. W. Trotter and M. R. Wood for discussion and critical reading of the manuscript; R. Cerione and K. Siminovitch for providing cDNAs for Cdc42 and WASP, respectively; L. Kay for providing many of the NMR pulse sequences used in this study; F. Delaglio for providing nmrPipe and TALOS software; A. Majumdar, W. Hu and B. Aghazadeh for assistance in NMR data acquisition; J. Hubbard for computer system support and M. Fiore for expert administrative assistance. A.S.K. is supported by a fellowship from the Damon Runyon-Walter Winchell Foundation, and L.T.K. by the Charles H. Revson Foundation. M.K.R. acknowledges support from the NIH (PECASE program), Arnold and Mabel Beckman Foundation, and Sidney Kimmel Foundation for Cancer Research.

Correspondence and requests for materials should be addressed to M.K.R. (e-mail: rosen@mrnmrl.ski.mskcc.org).

Discovery of calcium in Mercury's atmosphere

Thomas A. Bida*, Rosemary M. Killen† & Thomas H. Morgan‡

* W. M. Keck Observatory, CARA, 65-1120 Mamalahoa Highway, Kamuela, Hawaii 96743, USA

† Southwest Research Institute, 6220 Culebra Road, San Antonio, Texas 78228-0510, USA

‡ NASA Headquarters, 300 E. Street NW, Washington, District of Columbia 20546-1113, USA

The composition and evolutionary history of Mercury's crust are not well determined^{1,2}. The planet as a whole has been predicted³ to have a refractory, anhydrous composition: rich in Ca, Al, Mg and Fe, but poor in Na, K, OH, and S. Its atmosphere is believed to be derived in large part from the surface materials. A combination of effects that include impact vaporization (from infalling material), volatile evaporation, photon-stimulated desorption and sputtering releases material from the surface to form the atmosphere. Sodium and potassium have already been observed in Mercury's atmosphere^{4–6}, with abundances that require a volatile-rich crust⁷. The sodium probably results from photon-stimulated desorption^{8,9}, and has a temperature of 1,500 K (ref. 10). Here we report the discovery of calcium in the atmosphere near Mercury's poles. The column density is very low and the temperature is apparently very high (12,000 K). The localized distribution and high temperature, if confirmed, suggest that the atmospheric calcium may arise from surface sputtering by ions, which enter Mercury's auroral zone. The low abundance of atmospheric Ca may indicate that the regolith is rarefied in calcium.

We observed Mercury with the High Resolution Echelle Spectrograph (HIRES; ref. 11) at the W. M. Keck I telescope on the evenings of 7 July 1998, 13–19 July 1998 inclusive and on 21 July 1998 UT. Mercury was acquired at fixed sky position angles to orient a 7" long by 0.57" or 0.86" wide slit, tangent and perpendicular to both polar limbs, the sub-solar limb point, and the night-side of the disk. Each observed spectrum is a composite of narrow source-emission lines, an underlying planetary reflectance continuum, and terrestrial-sky continuum and emission. The exposure times varied from 100 to 600 s, with closed-loop guiding on Mercury and continuous pointing compensation for differential atmospheric dispersion. The results reported here were obtained in clear sky conditions, with good seeing (1.2") in very high airmass (up to 10).

Observations on 7 and 9 July revealed unambiguous Ca resonance emission from the CaI line (4,226.74 Å in air), as shown in Fig. 1. The prominent CaI emission lines observed on 7 and 19 July are Doppler-shifted from the solar continuum absorption core in proportion to the relative velocity of Mercury to the Sun (see Table 1). The emission spectra illustrated represent tangentially integrated column emission of 384 and 422 Rayleighs, sampled at a signal-to-noise ratio of 33 and 20 at line peak, for 7 and 19 July, respectively. Variable conditions on the other seven evenings of the 1998 campaign were responsible for a marginal detection of Ca emission on only two of those evenings (14 and 17 July).

The observed column emission intensity, emission linewidth, and residual Doppler-shift of Ca in Mercury's atmosphere are shown in Fig. 2 for three slit positions on two dates. The column emission decreased with altitude from Mercury's poles on 7 July, but is enhanced some 3,100 km W (local, that is, anti-sunward) of the south pole on 19 July. The line width is greater than instrumental broadening for all spatial samples (the average width of ThAr comparison lines in order 84 is 81 mÅ for 7 July, and 73 mÅ for 19 July; also see the line profiles in Fig. 1). The median line width of 130 mÅ observed on 7 July perpendicular to the south pole is

consistent with a Doppler temperature of 11,400 K; line widths plotted in Fig. 2 represent temperatures between 6,000 and 15,500 K. Localized enhancements in linewidth, which we relate to Ca temperature for comparison purposes, occur 2,100 km north of the north pole on 7 July, and 3,100 km W of the south pole on 19

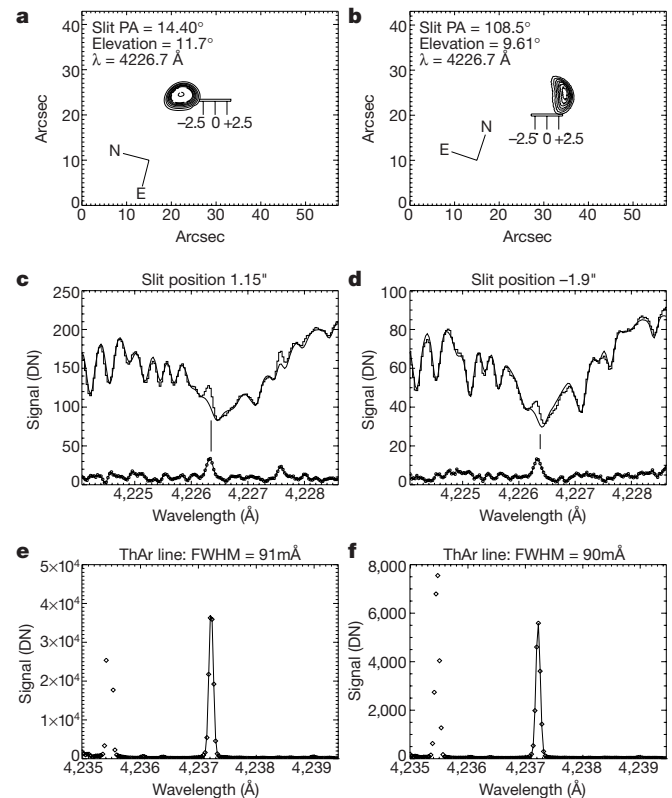


Figure 1 The observing geometry, samples of the CaI emission spectra, and the HIRES instrumental line profile. Data obtained for two observations on the evenings of 7 and 19 July 1998 UT. **a, c, e**, 7 July; **b, d, f**, 19 July. **a, b**, A contour map of a broadband acquisition image taken during the spectral exposures, with the slit position at Ca 4,226.7 Å. Positions along the slit are annotated in arcseconds relative to slit centre. PA is the position angle of the slit east of north. **c, d**, The extracted spectrum at one seeing-averaged slit position (histogram plot), the fitted sky continuum (solid overlay plot), and their difference, the emission spectrum (circle symbol plot). The expected position of the Doppler-shifted Ca emission line is indicated. On 7 July, a narrow scattered light feature contaminated the spectrum centred at 4,229.4 Å. The imaged spectra were reduced as follows. Each spectrum is bias, gain, and flat-field corrected, and the spatial profile of each echelle order defined by tracing a dispersed quartz continuum spectrum of a pinhole at slit centre. The two-dimensional orders are rectified, and the data is sampled according to seeing in a running mean along the slit, where the spatial plate scale is 0.19" per pixel. **e, f**, The spectrograph wavelength dispersion was determined with ThAr reference spectra taken immediately preceding the Mercury observations each evening. The 71- or 84-mÅ wide (2- or 3-pixel-wide slit deckers) line positions of catalogued ThAr lines within the order of interest (84, FSR 4,217–4,268 Å) are determined, and fitted with a sixth-order polynomial; the dispersion at CaI is 30 mÅ per pixel, and the r.m.s. deviations of the fits were less than 2 mÅ. Next, a twilight-sky spectrum obtained before sunset each evening is used to construct a two-component continuum spectrum, representing contributions from the planetary-solar reflectance and twilight-sky continua. The relative contributions of each component are iteratively scaled to minimize the r.m.s. deviation relative to each traced, averaged spectrum along the slit, and then subtracted from it; the twilight-sky contribution to the background was 63% for the 7 July spectrum shown, and less than 10% for the 19 July spectrum. The resulting emission spectrum is then analysed by gaussian line fitting for flux, linewidth, and line centre. The central wavelengths of the emission lines are then measured relative to polynomial fitted positions of the planetary-solar reflectance and twilight-sky line CaI absorption cores. The emission lines are expected to be Doppler-shifted from the absorption line cores because of the relative velocities of Mercury, the Sun and the Earth.

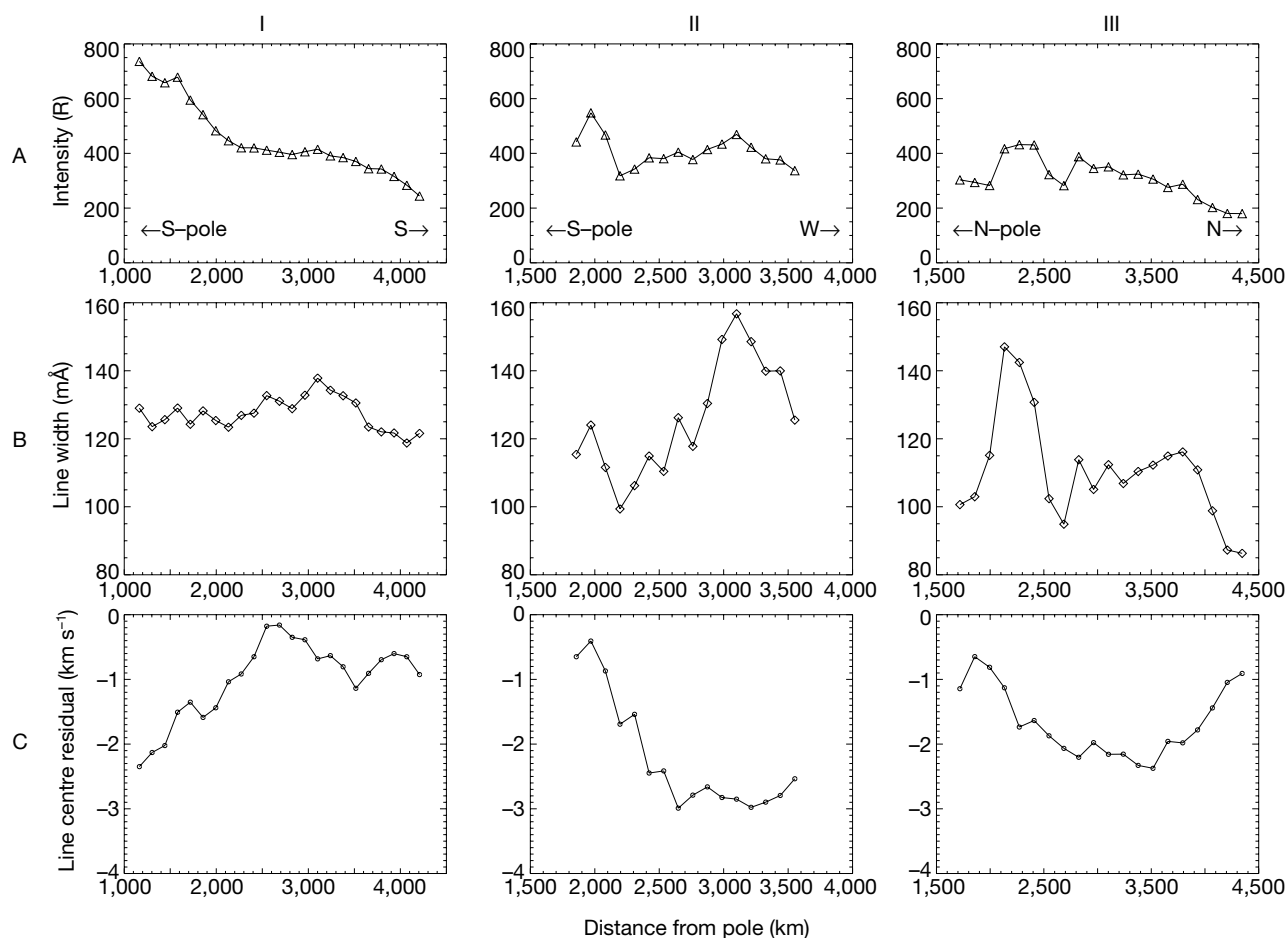


Figure 2 The spatial variation of the column emission intensity, emission line width, and residual Doppler-shift of Ca in Mercury's atmosphere. These measured quantities are shown where resolvable along the slit for three separate observations: 7 July 1998 south pole N–S direction (column I), 19 July 1998 south pole anti-solar (E–W) direction (column II), and 7 July 1998 north pole N–S direction (column III). In row A is plotted the CaI column

emission intensity (Rayleighs), row B the emission linewidth (mÅ), and row C the residual of the emission line centre (km s⁻¹) after subtraction of the Doppler shift due to the relative Earth–Mercury velocity (at -27.4 km s⁻¹, or 0.386 Å on 7 July). See the text for discussion of these plots.

July, where the column emission had also brightened. The most probable velocity for a Ca temperature of 12,000 K is 2.2 km s⁻¹, less than the escape velocity of 4.2 km s⁻¹. The Ca corona is apparently very dynamic and has velocity structure with altitude. The residual Doppler shift of the resonance line presented here is the line-of-sight (LOS) component of the Ca coronal velocity after subtraction of Mercury's velocity relative to the Earth; negative residuals represent a LOS component of motion towards Earth. On 7 July the LOS velocity was -2 km s⁻¹ near the south pole, decreasing to -0.2 km s⁻¹ 2,500 km due south, but gaining to -1 km s⁻¹ at 3,500 km altitude. Off the north pole 20 min later, the residual increased from -0.6 km s⁻¹ at 1,800 km to -2.1 km s⁻¹ at 3,000 km altitude, before falling off in magnitude. On 19 July, the steep gradient in the LOS velocity magnitude with position anti-sunward of the south pole, from -0.3 to -3 km s⁻¹ over 800 km, and the orbital geometry ($\varphi = 101^\circ$ (Table 1)) are such that the Ca velocity is increasing rapidly while moving anti-sunward of the pole at that time. All of these observations describe a Ca corona that varies both spatially and temporally. We found no evidence of CaI emission in the vicinity of the sub-solar limb; however, any emission there may have been obscured by high surface reflectance. An off-pole detection is favoured as a result of reduced continuum and long LOS.

The observational parameters, observed brightness, and derived Ca abundances for the select emission spectra of Fig. 1 are given in Table 1. The tangentially integrated density, N_{tangent} , represents

integration over the LOS and is obtained from the column emission intensity $4\pi I$ as (ref. 12, equation 6.2.3):

$$N_{\text{tangent}} = 4\pi I/g \quad (1)$$

The g value (in s⁻¹) is the photon-scattering probability from Ca at Mercury orbit, given by:

$$g = (\pi F_s / r_m^2) (\pi e^2 / mc) f (\lambda^2 / c) \quad (2)$$

Here the solar flux at Earth orbit is πF_s (in photons cm⁻² s⁻¹ Å⁻¹; ref. 13). (The solar flux at 423 nm at Earth orbit is 239 ergs cm⁻² s⁻¹ Å⁻¹, or 5.08×10^{13} photons cm⁻² s⁻¹ Å⁻¹.) r_m is the orbital distance of Mercury (in AU); the f -value, or oscillator strength, is 1.75 for the multiplet; and $\pi e^2 / mc$ is the integrated absorption coefficient per atom per f -value. If the exosphere is not spherically symmetric, it is reasonable to expect that $N_{\text{zenith}} \approx N_{\text{tangent}}$. The average measured zenith column abundance is then 1.1×10^8 cm⁻² for these two samples. By comparison, the Na zenith abundance is $\sim 2 \times 10^{11}$ cm⁻².

There are three possible source processes for neutral Ca in Mercury's atmosphere: photon-stimulated desorption, impact vaporization from infalling meteoritic material, and ionic sputtering¹⁴. The large (>100 mÅ) Doppler contribution to the Ca linewidth equates to a temperature of about 12,000 K in a thermalized atmosphere. Exospheric linewidths are expected to be narrower than their atmospheric counterparts with the same mean

Table 1 Observational parameters

Date	Z (R _m)	4πI (R)	v (km s ⁻¹)	φ (°)	γ	g (s ⁻¹)	N _{tangent} (cm ⁻²)
7 July 1998	2.25	384	7.561	79	0.14	10.46	3.67 × 10 ⁷
19 July 1998	1.4	422	2.249	101	0.0352	2.29	1.8 × 10 ⁸

The observational parameters, emission intensity, and derived Ca abundance for the selected emission spectra of Fig. 1. The radial distance, Z, is measured from the planet's centre to the tangent point of the line-of-sight to the observer in units of planetary radii (R_m); 4πI is the column emission intensity in Rayleighs (1 R = 10⁶ photons cm⁻² s⁻¹); v is the velocity of the planet radial to the sun; φ is the Sun–Mercury–Earth phase angle; γ is the fraction of solar continuum at the line centre wavelength as seen by atoms at rest in Mercury's atmosphere; g is the scattering coefficient described in the text; and N_{tangent} is the column abundance tangent to the line of sight.

energy^{15,16}; therefore, the assignment of an equivalent temperature of 12,000 K probably underestimates the mean energy of the gas. This high temperature is expected for an ion-sputtered source¹⁷, but is at least twice that expected from an impact vapour source¹⁸, and ten times that expected from a photo-sputtered source¹⁵. The localization of Ca in the vicinity of the polar regions, if verified by subsequent observations, along with increased column abundance in the anti-solar direction, supports the hypothesis that the Ca in the atmosphere is sputtered by ions directed poleward by Mercury's magnetosphere. We note that the Ca–O bond strength is 4.8 eV, in contrast with 2.6 eV for the Na–O bond, so that Ca may be sputtered in a larger proportion bound as Ca–O than is the case for Na and Na–O.

The low-Ca zenith abundance (1.1 × 10⁸ cm⁻²) is much less than previous predicted (6.4 × 10⁸ cm⁻², ref. 19) and observational upper limits (7.4 × 10⁸ cm⁻², ref. 20). Although the refractory nature of Ca will limit its supply to the atmosphere, the observed low density and apparent polar concentration of Ca are consistent with a surface composition that is more volatile-rich than that of the Moon¹⁹. The recent identification of labradorite feldspar, a 30%–50% Na-rich albite, on Mercury²¹, along with other infrared observations²², is additional evidence for a Ca-poor material. Alternatively, there may be patchy locations of Na-rich, K-rich (Ca-poor) materials²³. □

Received 23 July; accepted 15 December 1999.

- Goettel, K. A. in *Mercury* 613–621 (Univ. Arizona Press, Tucson, 1988).
- Lewis, J. S. in *Mercury* 651–666 (Univ. Arizona Press, Tucson, 1988).
- Lewis, J. S. Metal/silicate fractionation in the solar system. *Earth Planet Sci. Lett.* **15**, 286–290 (1992).
- Potter, A. E. & Morgan, T. H. Discovery of sodium in the atmosphere of Mercury. *Science* **229**, 651–653 (1985).
- Potter, A. E. & Morgan, T. H. Potassium in the atmosphere of Mercury. *Icarus* **67**, 336–340 (1986).
- Sprague, A. L. *et al.* Distribution and abundance of sodium in Mercury's atmosphere, 1985–1988. *Icarus* **129**, 506–527 (1997).
- Killen, R. M. & Morgan, T. H. Maintaining the Na atmosphere of Mercury. *Icarus* **101**, 293–312 (1993).
- McGrath, M. A., Johnson, R. E. & Lanzerotti, L. J. Sputtering of sodium on the planet Mercury. *Nature* **323**, 694–696 (1986).
- Yakshinskiy, B. V. & Madey, T. E. Photon-stimulated desorption as a substantial source of sodium in the lunar atmosphere. *Nature* **400**, 642–644 (1999).
- Killen, R. M., Potter, A. E., Fitzsimmons, A. & Morgan, T. H. Sodium D2 line profiles: clues to the temperature structure of Mercury's exosphere. *Planet. Space Sci.* **47**, 1449–1458 (1999).
- Vogt, S. S. *et al.* HIREs: the high-resolution echelle spectrometer on the Keck 10-m telescope. *Proc. SPIE* **2198**, 362 (1994).
- Chamberlain, J. W. & Hunt, D. M. *Theory of Planetary Atmospheres* (Academic Press, Orlando, 1987).
- Kurucz, R. L., Furenlid, I., Brault, J. & Testerman, L. *Solar Flux Atlas From 296 to 1300 nm*, National Solar Observatory Atlas No. 1 (Harvard Univ. Press, Boston, 1984).
- Cheng, A. F., Johnson, R. E., Krimigis, S. M. & Lanzerotti, L. J. Magnetosphere, exosphere and surface of Mercury. *Icarus* **71**, 430–440 (1987).
- Madey, T. E., Yakshinskiy, B. V., Ageev, V. N. & Johnson, R. E. Desorption of alkali atoms and ions from oxide surfaces: relevance to origins of Na and K atmospheres of Mercury and the Moon. *J. Geophys. Res.* **103**, 5873–5887 (1998).
- Chamberlain, J. W. Spectral line profiles for a planetary corona. *J. Geophys. Res.* **81**, 1774–1776 (1976).
- Elphic, R. C. *et al.* Lunar surface composition and solar wind-induced secondary ion mass spectrometry. *Geophys. Res. Lett.* **18**, 2165–2168 (1991).
- Sugita, S., Schultz, P. H. & Adams, M. A. In situ temperature measurements of impact-induced vapor clouds with spectroscopic methods. *Proc. Lunar Planet. Sci. Conf.* **28**, 1393–1394 (1997).
- Morgan, T. H. & Killen, R. M. A non-stoichiometric model of the composition of the atmospheres of Mercury and the Moon. *Planet. Space Sci.* **45**, 81–94 (1997).
- Sprague, A. L., Koslowski, R. W. H., Hunt, D. M. & Grosse, F. A. An upper limit on neutral calcium in Mercury's atmosphere. *Icarus* **104**, 33–37 (1993).
- Sprague, A. L., Nash, D. B., Witteborn, F. C. & Cruikshank, D. P. Mercury's feldspar connection: mid-IR measurements suggest plagioclase. *Adv. Space Res.* **19**, 1507–1510 (1997).

- Jeanloz, R., Mitchell, D. L., Sprague, A. L. & dePater, I. Evidence for a basalt free surface on Mercury and implications for internal heat. *Science* **268**, 1455–1457 (1995).
- Sprague, A. L., Schmitt, W. J. & Hill, R. E. Mercury: Sodium atmospheric enhancements, radar bright spots and surface features. *Icarus* **136**, 60–68 (1998).

Acknowledgements

The data presented here were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership between the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The observatory was made possible by the generous financial support of the W.M. Keck Foundation. We are grateful to the WMKO Observing Assistants for their help while conducting the observations. R.M.K. acknowledges the support of a NASA grant.

Correspondence and requests for materials should be addressed to T.A.B. (e-mail: tbida@keck.hawaii.edu).

North–south geological differences between the residual polar caps on Mars

P. C. Thomas^{*}, M. C. Malin[†], K. S. Edgett[‡], M. H. Carr[§], W. K. Hartmann[§], A. P. Ingersoll^{||}, P. B. James[¶], L. A. Soderblom^{**}, J. Veverka^{*} & R. Sullivan^{*}

^{*} Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA

[†] Malin Space Science Systems, P.O. Box 910148, San Diego, California 92191, USA

[‡] US Geological Survey, Menlo Park, California 94025, USA

[§] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA

[¶] Planetary Science Institute, Tucson, Arizona 85719, USA

^{||} Department of Physics and Astronomy, University of Toledo, Toledo, Ohio 43606, USA

^{**} US Geological Survey, Flagstaff, Arizona 86001, USA

Polar processes can be sensitive indicators of global climate, and the geological features associated with polar ice caps can therefore indicate evolution of climate with time. The polar regions on Mars have distinctive morphologic and climatologic features: thick layered deposits, seasonal CO₂ frost caps extending to mid latitudes, and near-polar residual frost deposits that survive the summer^{1,2}. The relationship of the seasonal and residual frost caps to the layered deposits has been poorly constrained^{3,4}, mainly by the limited spatial resolution of the available data. In particular, it has not been known if the residual caps represent simple thin frost cover or substantial geologic features. Here we show that the residual cap on the south pole is a distinct geologic unit with striking collapse and erosional topography; this is very different from the residual cap on the north pole, which grades into the underlying layered materials. These findings indicate that the differences between the caps are substantial (rather than reflecting short-lived differences in frost cover), and so support the idea of long-term asymmetry in the polar climates of Mars.

Data from the Mariner 9 and Viking missions established the distribution of the three basic polar units on Mars. Layered deposits centred approximately on the poles, consisting of unknown proportions of water ice and dust averaging 1,200 km across and up to 3 km in thickness⁵, seasonal carbon dioxide (CO₂) frost caps that remove about 25% of the martian atmosphere, and residual summer caps of frost: water ice in the north, CO₂ in the south^{6–8}. The northern residual cap is nearly coextensive with the layered deposits; the southern one is very much smaller in extent than the layered deposits. Mariner 9 and Viking images could not resolve the relation of the residual frost caps to the underlying layered deposits.

Images from the Mars Orbiter Camera (MOC) on the Mars Global Surveyor spacecraft, with pixel scales as good as 1.5 m (ref. 9; 30–50 times better than most previous images) now provide comparisons of the residual caps and layers in both polar deposits.

Images from the MOC show that the north-polar residual cap surface has a rough ablational topography of pits, cracks and some knobs (Fig. 1). The residual cap surface grades into exposures of layers (each typically 1–3 m thick) at the edges of dark lanes that expose the underlying layered deposits in low-slope cross-section^{10,11} (Fig. 1c). Layers exposed in the dark lanes are commonly expressed as subtle ridges or troughs¹² on the slopes, and usually have a disaggregated appearance.

In contrast to the north, the south residual cap has distinctive erosional or collapse topography (Fig. 2) affecting four or more layers, each about two metres thick (determined from shadow measurements). The top surface of the upper layer has polygonal depressions suggestive of thermal contraction cracks (Fig. 2a). Circular and cycloidal depressions, individual troughs with varying curvatures and groups of nearly parallel troughs giving ‘fingerprint’ patterns are also observed.

These distinctive morphologies occur throughout, but only within the south residual cap as mapped by Viking in 1977 (ref. 13). Isolated examples of degraded depressions and ridges are found in a few locations outside the residual cap in the south-polar layered deposits, but are not readily assigned to versions of the forms seen on the residual cap. These areas of the layered deposits have been thermally modelled¹⁴ as dust-covered and mechanically distinct from the residual cap. Exposed layers or unconformities in polar deposits in both hemispheres do not show the kind of relief found on the southern residual cap.

Thicknesses of layers in the south residual cap area are not much greater than that predicted for seasonal CO₂ deposition near the pole (0.5–1 m; refs 1 and 8); this similarity raises the possibility that the features might be short-lived. Images at 1.5 to 5 metres per pixel

acquired over five (Earth) months of the southern spring show no changes in the topography. Dark markings observed by Mariner 9 in 1972 at 120 metres per pixel correlate with arcuate remnants of the upper layer imaged by MOC. The persistence of a few metres of relief, less than a kilometre wide, for 27 years (scarp retreat of less than a few metres per year) in an area where material has been removed over horizontal scales of tens of kilometres, suggests that these distinctive forms may have survived for more than 1,000 years. Patterned materials at the bases of scarps (not shown here) may suggest long periods of thermal cycling¹⁵, which at polar latitudes is nearly all seasonal; thus modification of the aprons alone might span hundreds of years. There are no obvious impact craters with diameters greater than 100 m in the 550+ MOC images of the south residual cap with pixel scales less than 20 m. These images cover about 35% of ~48,000 km² of the residual cap accessible to MOC (to 87° S). The lack of craters suggests a crater retention age of less than 10⁵ years, substantially less than the more than 10⁷ years inferred for the southern layered deposits outside the residual cap^{16,17}.

The MOC images of the southern residual cap also indicate that coverage by the seasonal cap is either thin, nearly transparent⁸, or both. Although there might be a small component similar to the CO₂ ice ‘slabs’ inferred for other parts of the seasonal cap⁸, the upper surface of the residual cap has a very high albedo in the spring and shows albedo features indicative of differential defrosting of some areas. However, there are no thick accumulations of drifted snow, and the crispness of the topography viewed with solar elevation angles of only two degrees suggests less than a metre of seasonal covering within the residual cap area.

Images from MOC indicate the following sequence of events in the south-polar residual cap area: (1) Deposition of at least the four upper layers on a lower surface. (2) Sag in areas a few tens of metres to a few hundred metres across in a large fraction of the residual cap area. (3) Collapse over sags and lateral expansion of depressions,

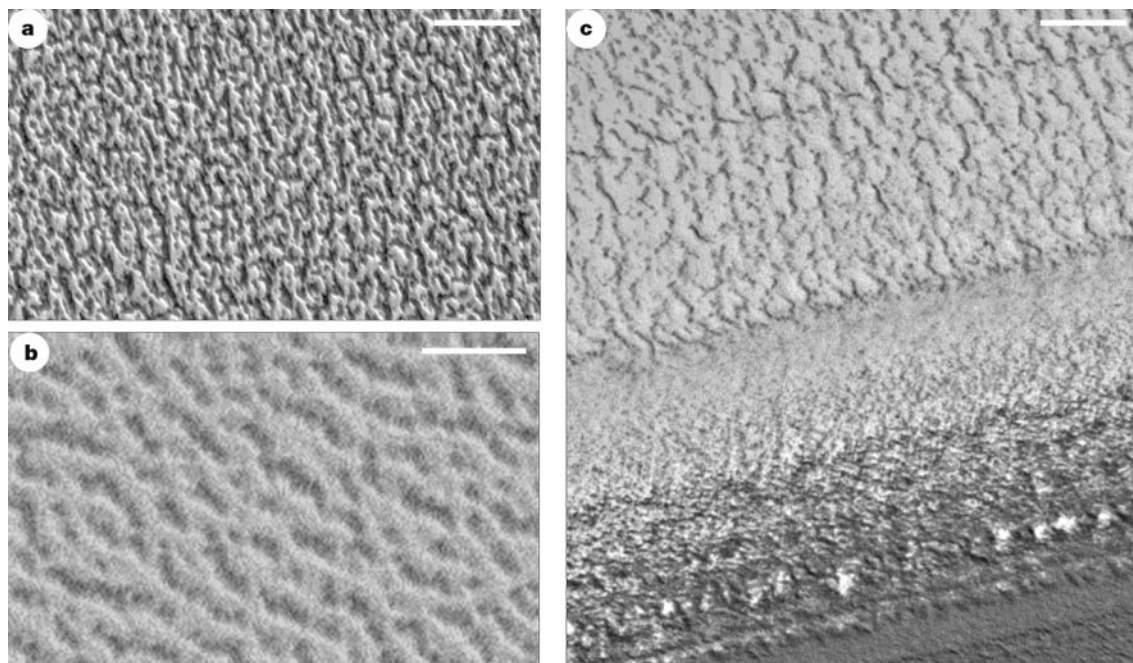


Figure 1 North-polar residual cap topography. **a**, Individual and connected pits in surface of residual cap. Portion of Mars Orbiter Camera (MOC) image M00-00547 (mapping phase 0, image 547), 82.1° N, 329.6° W, solar incidence angle $i = 73^\circ$. Texture on most of the north-polar residual cap is a variant of pitting of approximately similar width depressions; length and connectivity of depressions varies. Depths inferred from minimal evidence of shadows are probably less than 2 m. Scale bar is 200 m. Illumination is from upper right. Aerocentric longitude of the sun (L_s) is 120° (0° is northern spring equinox). 5 April 1999. This and other images shown have been contrast enhanced. **b**, North-polar

residual cap surface, portion of image CAL-00433; 86.9° N, 207.5° W; scale bar is 50 m. Illumination is from upper right. $L_s = 107^\circ$. 8 March 1999. **c**, Merging of pits and fissures of residual cap topography with exposure of layers in walls of one of the dark lanes (troughs with wall slopes generally less than 10°) that traverse much of the northern layered deposits; area at bottom of image is largely frost-free. Portion of MOC image M00-02072; 85.9° N, 258.1° W; $i = 70^\circ$. Scale bar is 100 m. Illumination is from upper right. $L_s = 124^\circ$. 13 April 1999.

perhaps aided by insolation effects at the steeper parts of collapse depressions. (4) Removal of large areas of volatile materials leaving lag surfaces with less than 2 m of relief. (5) Formation of debris aprons around higher (9 m) scarps. (6) Production of moats around higher residual areas either by deposition of 1–2 m of material around debris aprons followed by loss of some apron material, or by further compaction of underlying materials by debris aprons and subsequent loss of some material from the debris aprons. (7) Further development of debris aprons, with a horizontal change of less than 20 m in scarp positions. Time scales associated with each specific step are unknown, although as noted above, the total time could greatly exceed 1,000 years.

Important features in the above sequence of events are the initial generation and form of collapse (sags), depression growth (possibly by sublimation), removal of large amounts of material (8 m by several thousand square kilometres) and a possible additional depositional period.

What causes the initial sags and collapse? The volume change at depth could be due to compaction or to removal of material. Compaction of layers that were anomalously porous might occur

with further deposition in combination with heating owing to rising isotherms with increasing time of burial. The plan form of compacting material could reflect variations in thickness or texture of the original deposit, such as snow dunes. Removal of material would most probably be caused by sublimation of CO₂ or water ice. In either case, the collapse regime is apparently restricted to the area currently showing summer residual CO₂ at the surface. The circumstances suggest that CO₂ rather than water constitutes the collapsing unit. Carbon dioxide is the most likely constituent that could sublimate to initiate the sags because of its much higher vapour pressure and because of the apparent lack of collapse in areas of the water-ice-covered north-polar residual cap, or in the outlying south-polar layered deposits. That is, if water ice is the chief volatile constituent of the polar deposits¹¹, CO₂ deposits in the residual cap area would provide the contrast in volatility that the surface morphology suggests.

Although thick accumulations of CO₂ ice are probably not possible¹⁸, a few metres of coverage could occur as they would have minimal basal pressure (0.6 MPa for 10 m of solid CO₂ ice) and would extract a very small fraction of the atmospheric CO₂

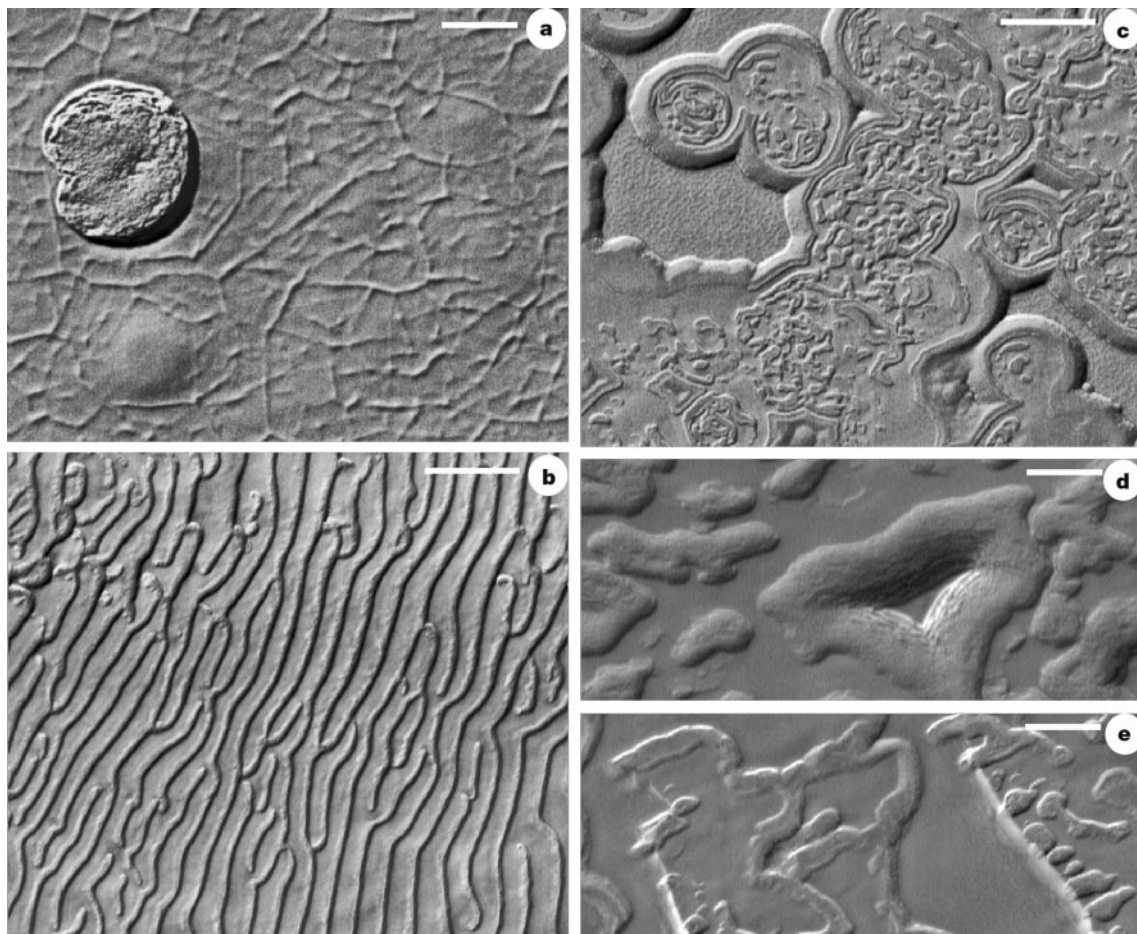


Figure 2 South-polar residual cap topography. **a**, Nearly circular depression and nearby sag surface on top layer. Polygonal cracks are prominent on undisturbed sections of the upper layer surface, but do not affect the margins of circular depressions. Portion of MOC image M09-00609; 87.0° S, 5.9° W; $i = 70^\circ$. Scale bar is 100 m. Illumination is from lower right. $L_s = 237^\circ$. 3 November 1999. **b**, 'Fingerprint' pattern of depressions. Elongated sags in other areas of residual cap suggest precursors of this topography. Steeper sides (right) of the troughs face in a more northerly direction, suggesting sublimation in expanding depressions, as with the more circular ones. Portion of MOC image M03-06756; 86.0° S, 53.9° W; $i = 88^\circ$. Illumination is from lower right. Scale bar is 500 m. $L_s = 182^\circ$. 4 August 1999. **c**, Circular collapse features, leaving mesas on upper surface with debris aprons and moats. Largest scarps here are about 4 m high. The uppermost layer is capable of supporting scarp slopes of $\sim 20^\circ$; the aprons frequently

have slopes of order 1° – 3° (Fig. 2b); slopes are estimated from presence or absence of shadows as the sun gained elevation in the southern spring. Portion of MOC image M03-06646; 85.6° S, 74.4° W; $i = 88^\circ$. Scale bar is 500 m. Illumination is from lower right. $L_s = 181^\circ$. 3 August 1999. **d**, Residual mesa exposing four layers and surrounding moat. Formation of moats probably requires additional deposition and sublimation or compaction following removal of material from height of top layer exposed here. Portion of MOC image M07-02129; 86.9° S, 78.5° W; $i = 81^\circ$. Scale bar is 100 m. Illumination is from bottom, right. $L_s = 204^\circ$. 11 September 1999. **e**, Complex covering of depressions which suggest burial and exhumation of topography on part of the southern residual cap area. Portion of MOC image M04-03877; 84.6° S, 45.1° W, $i = 81^\circ$, scale bar is 200 m. Illumination is from lower right. $L_s = 196^\circ$. 29 August 1999.

reservoir. The removal of wide areas of the upper 8–10 m of the residual cap (Fig. 2c and d) after initial sag and collapse implies that most of those upper layers are also subject to sublimation, especially once the upper surface is fragmented or disturbed. Most of the eight metres that are removed is probably CO₂; the lag might be less volatile water ice. Radar data returns from the south-polar residual cap are very different from the north-polar cap¹⁹, and can be interpreted as indicative of clean ice with large cavities.

The evolution of the topography after collapse (steps 4–7 above) suggests at least two subsequent non-steady-state periods. The wide area of removal of three or more layers (Fig. 2c and d) indicates considerable sublimation. The circular and near-circular forms may develop because at these high latitudes, ablational removal can become nearly azimuthally isotropic owing to the small change in solar elevation during the day. The remnant landforms suggest lag deposits from ablation and further collapse. The small proportion remaining as lag (Fig. 2a and c) suggests a small non-volatile (or less volatile) component in the upper layers of the residual cap area. The moats (Fig. 2c and d) also require some sequence of renewed deposition and subsequent partial removal, or a period of debris apron growth followed by apron retreat or compression of underlying layers. Additionally, some areas show probable burial and partial exhumation of the depressions (Fig. 2e).

The MOC data show that the southern residual cap is not simply a temporary anomaly of residual summer CO₂ frost; it is a geological feature indicative of depositional and ablational events recorded neither in the north nor in the outlying southern polar layered deposits. The geographical restriction to the CO₂ residual cap strongly suggests that CO₂ ice is involved, as does the apparent requirement for a component distinct from water ice. The most obvious environmental distinctions of the southern residual cap area are its elevation, about 6 km above the northern one^{11,20}, and its presence in the hemisphere that has very low atmospheric water content².

Indications of burial and exhumation (Fig. 2e, and steps 6 and 7 above) suggest repetitive, or even periodic formation of the distinctive southern residual cap morphology. Change in hemispheric asymmetry in polar processes has been attributed to periodic variations^{21–24} in the orbital eccentricity, obliquity and season of perihelion of Mars. Even the shortest of these cycles, 51,000 years for season of perihelion, could easily allow the build-up (or sublimation) of the whole of the south residual cap unit. Eight metres of solid CO₂ accumulated as net residual in 1,000 years would require only about 2% of the current seasonal total¹⁸ CO₂ to be retained each year. However, elucidation of the time scale represented by the residual deposits rests on detection of the net atmospheric CO₂ budget and on mapping cycles recorded in the layered deposits as a whole. □

Received 3 December 1999; accepted 7 January 2000.

- James, P. B., Kieffer, H. H. & Paige, D. A. in *Mars* (eds Kieffer H., Jakosky, B., Snyder C. & Matthews, M.) 934–968 (Univ. Arizona Press, Tucson, 1992).
- Jakosky, B. M. & Haberle, R. M. in *Mars* (eds Kieffer H., Jakosky, B., Snyder C. & Matthews, M.) 969–1016 (Univ. Arizona Press, Tucson, 1992).
- Pollack, J. B., Colburn, D., Flasar, F. M., Kahn, R., Carlston, C. & Pidek, D. Properties and effects of dust particles suspended in the Martian atmosphere. *J. Geophys. Res.* **84**, 2929–2945 (1979).
- Thomas, P. C., Herkenhoff, K., Howard, A. & Murray, B. in *Mars* (eds H. Kieffer, H., Jakosky, B., Snyder, C. & M. Matthews) 767–795 (Univ. Arizona Press, Tucson, 1992).
- Smith, D. E. *et al.* The global topography of Mars and implications for surface evolution. *Science* **284**, 1495–1503 (1999).
- Farmer, C. B., Davies, D. W. & LaPorte, D. D. Mars: Northern summer ice cap—water vapor observations from Viking 2. *Science* **194**, 1339–1340 (1976).
- Kieffer, H. H. Mars south polar spring and summer temperatures. A residual CO₂ frost. *J. Geophys. Res.* **84**, 8263–8288 (1979).
- Kieffer, H. H., Titus, T. N., Mullins, K. F. & Christiansen, P. R. Mars south polar spring and summer behavior observed by TES; seasonal cap evolution controlled by grain size. *J. Geophys. Res.* (submitted).
- Malin, M. C. *et al.* Early views of the Martian surface from the Mars Orbiter Camera of Mars Global Surveyor. *Science* **279**, 1681–1685 (1998).
- Howard, A. D., Cutts, J. A. & Blasius, K. R. Stratigraphic relationships within Martian polar cap deposits. *Icarus* **50**, 161–215 (1982).
- Zuber, M. T. *et al.* Observations of the north polar region of Mars from the Mars Orbiter Laser Altimeter. *Science* **282**, 2053–2060 (1998).

- Malin, M. C. & Edgett, K. S. The nature of layered outcrop expression in the martian polar layered terrains. *Lunar Planet. Sci.* **34**, Abstr. no. 1055 (2000).
- James, P. B., Briggs, G., Barnes J. & Spruck, A. Seasonal recession of Mars' south polar cap as seen by Viking. *J. Geophys. Res.* **84**, 2889–2922 (1979).
- Paige, D. A. & Keegan, K. D. Thermal and albedo mapping of the polar regions of Mars using Viking thermal mapper observations: 2. South polar region. *J. Geophys. Res.* **99**, 25993–26031 (1994).
- Mellon, M. T. Small-scale polygonal features on Mars: Seasonal thermal contraction cracks in permafrost. *J. Geophys. Res.* **102**, 25617–25628 (1997).
- Plaut, J. J., Kahn, R., Guinness, E. A. & Arvidson, R. E. Accumulation of sedimentary debris in the south polar region of Mars and implications for climate history. *Icarus* **76**, 357–377 (1988).
- Herkenhoff, K. E. & Plaut, J. J. Surface ages and resurfacing rates of the polar layered deposits on Mars. *Icarus* (in the press).
- Mellon, M. T. Limits on the CO₂ content of the martian polar deposits. *Icarus* **124**, 268–279 (1996).
- Muhleman, D. O., Grossman, A. W. & Butler, B. J. Radar investigation of Mars, Mercury, and Titan. *Annu. Rev. Earth Planet. Sci.* **23**, 337–374 (1995).
- Schenk, P. M. & Moore, J. M. Stereo topography of the south polar region of Mars: Volatile inventory and Mars landing site. *J. Geophys. Res.* (in the press).
- Toon, O. B., Pollack, J. B., Ward, W., Burns, J. A. & Bilski, K. The astronomical theory of climatic change on Mars. *Icarus* **44**, 552–607 (1980).
- Kieffer, H. H. & Zent, A. P. in *Mars* (eds H. Kieffer, H., Jakosky, B., Snyder, C. & M. Matthews) 1180–1220 (Univ. Arizona Press, Tucson, 1992).
- Ward, W. R. in *Mars* (eds Kieffer, H. H., Jakosky, B., Snyder, C. & Matthews, M. S) 298–320 (Univ. Arizona Press, Tucson 1992).
- Cutts, J. A. & Lewis, B. H. Models of climate cycles recorded in Martian polar layered deposits. *Icarus* **50**, 216–244 (1982).

Acknowledgements

E. Jensen, M. Roth, M. Ryan, D. Sharman and J. Warren provided technical assistance. This work was supported in part by NASA.

Correspondence and requests for materials should be addressed to P.C.T. (e-mail: thomas@cusplif.tn.cornell.edu).

Impossibility of deleting an unknown quantum state

Arun Kumar Pati*† & Samuel L. Braunstein*

* Quantum Optics and Information Group, Informatics, Dean Street, University of Wales, Bangor LL57 1UT, UK

† Theoretical Physics Division, BARC, Bombay 400 085, India

A photon in an arbitrary polarization state cannot be cloned perfectly^{1,2}. But suppose that at our disposal we have several copies of a photon in an unknown state. Is it possible to delete the information content of one or more of these photons by a physical process? Specifically, if two photons are in the same initial polarization state, is there a mechanism that produces one photon in the same initial state and the other in some standard polarization state? If this could be done, then one would create a standard blank state onto which one could copy an unknown state approximately, by deterministic cloning^{3,4} or exactly, by probabilistic cloning^{5,6}. This could in principle be useful in quantum computation, where one could store new information in an already computed state by deleting the old information. Here we show, however, that the linearity of quantum theory does not allow us to delete a copy of an arbitrary quantum state perfectly. Though in a classical computer information can be deleted (reversibly) against a copy⁷, the analogous task cannot be accomplished, even irreversibly, with quantum information.

Quantum information has the unique property that it cannot be amplified accurately. If an arbitrary state could be cloned, then by using non-local resources one could send signals faster than light^{1,2}. However, orthogonal quantum states can be perfectly copied. Although two non-orthogonal photon-states cannot be copied perfectly by a unitary process⁸, they can be copied by a unitary-reduction process⁵. More interestingly, non-orthogonal states

from a linearly independent set can evolve into a linear superposition of multiple copy states using a new cloning machine⁹. With the recent advances in quantum information theory, such as quantum cryptography¹⁰, quantum teleportation^{11–14} and quantum computing¹⁵, we would like to know what we can do with the vast amount of information contained in an unknown state and what we cannot.

As is now well understood, erasing a single copy of some classical information is an irreversible operation. It may only be done with some energy cost; a result known as the Landauer erasure principle⁷. In quantum theory, erasure of a single unknown state may be thought of as swapping it with some standard state and then dumping it into the environment. The deletion we wish to study is not the same as irreversible erasure; it is more like reversible ‘uncopying’ of an unknown quantum state. We will show that there is no quantum deleting machine that can delete one unknown state against a copy in either a reversible or an irreversible manner.

Let us suppose that we have several copies of some unknown information. Classically we may delete one copy against the others, uncopying it in a perfectly reversible manner⁷. The situation is very different in quantum theory. Such a quantum deleting machine would involve two initially identical qubits (for example, photons of arbitrary polarization) in some state $|\psi\rangle$ and an ancilla in some initial state $|A\rangle$, which corresponds to the ‘ready’ state of the deleting apparatus. The aim of this machine is to delete one of two copies of $|\psi\rangle$ and replace it with some standard state of a qubit $|\Sigma\rangle$. The quantum deleting operation is defined for an input $|\psi\rangle|\psi\rangle$ such that the linear operator acts on the combined Hilbert space of input and ancilla. That is, it is defined by:

$$|\psi\rangle|\psi\rangle|A\rangle \rightarrow |\psi\rangle|\Sigma\rangle|A_\psi\rangle \quad (1)$$

Here $|A_\psi\rangle$ is the final state of the ancilla, which may in general depend on the polarization of the original photon. (If we knew that this process was unitary, it might work like the time-reverse of cloning.) One obvious solution to this equation is to swap the second and third states. However, this reduces to the standard erasure result where the extra copies have played no role. We will therefore explicitly exclude swapping as describing quantum deleting.

Consider the action of this deleting machine, equation (1), on a pair of horizontally or vertically polarized photons:

$$\begin{aligned} |H\rangle|H\rangle|A\rangle &\rightarrow |H\rangle|\Sigma\rangle|A_H\rangle \\ |V\rangle|V\rangle|A\rangle &\rightarrow |V\rangle|\Sigma\rangle|A_V\rangle \end{aligned} \quad (2)$$

We note that the transformation defining our deleting machine, equation (1), does not completely specify its action when the input states are non-identical. This is in contrast to the Wootters–Zurek cloning transformation¹, whose definition specifies its action for all possible inputs. Because of this the transformation corresponding to our machine is not the time-reverse of cloning. In fact, the transformation (1), defines a whole class of possible deleting machines which could behave differently if the two inputs are unequal or even entangled, for example,

$$\frac{1}{\sqrt{2}}(|H\rangle|V\rangle + |V\rangle|H\rangle)|A\rangle \rightarrow |\Phi\rangle \quad (3)$$

where $|\Phi\rangle$ might be any state of the combined input–ancilla system.

Now for an arbitrary input qubit $|\psi\rangle = \alpha|H\rangle + \beta|V\rangle$ (where α and β are unknown complex numbers with $|\alpha|^2 + |\beta|^2 = 1$), linearity and the transformations (2) and (3) show that the deleting machine yields

$$\begin{aligned} &|\psi\rangle|\psi\rangle|A\rangle \\ &= [\alpha^2|H\rangle|H\rangle + \beta^2|V\rangle|V\rangle + \alpha\beta(|H\rangle|V\rangle + |V\rangle|H\rangle)]|A\rangle \\ &\rightarrow \alpha^2|H\rangle|\Sigma\rangle|A_H\rangle + \beta^2|V\rangle|\Sigma\rangle|A_V\rangle + \sqrt{2}\alpha\beta|\Phi\rangle \end{aligned} \quad (4)$$

which is a quadratic polynomial in α and β . However, if transformation (1) is to hold, transformation (4) must reduce to

$$(\alpha|H\rangle + \beta|V\rangle)|\Sigma\rangle|A_\psi\rangle \quad (5)$$

for all α and β . As $|\Phi\rangle$ is independent of α and β then $|A_\psi\rangle$ must be linear in α and β with the only solutions being $|\Phi\rangle = (|H\rangle|\Sigma\rangle|A_V\rangle + |V\rangle|\Sigma\rangle|A_H\rangle)/\sqrt{2}$ and $|A_\psi\rangle = \alpha|A_H\rangle + \beta|A_V\rangle$. Further, since the final state (5) must be normalized for all possible α and β , it follows that the ancilla states $|A_H\rangle$ and $|A_V\rangle$ are orthogonal. However, as discussed above, transformation (1) is therefore not uncopying at all, but merely swapping onto a two-dimensional subspace of the ancilla. It appears that there is no option but to move the information around without deleting it. That is, the linearity of quantum theory forbids deleting one unknown state against a copy. This we call the ‘quantum no-deleting’ principle. This principle is complementary in spirit to the no-cloning principle, and we expect it to play a fundamental role in future understanding of quantum information theory.

We emphasize that copying and deleting of information in a classical computer are inevitable operations whereas similar operations cannot be realized perfectly in quantum computers. This may have potential applications in information processing because it provides intrinsic security to quantum files in a quantum computer. No one can obliterate a copy of an unknown file from a collection of several copies in a quantum computer. In spite of the quantum no-deleting principle one might try to construct a universal and optimal approximate quantum deleting machine by analogy with optimal quantum cloning machines¹⁶. When memory in a quantum computer is scarce (at least for a finite number of qubits), approximate deleting may play an important role in its own way. Although at first glance quantum deleting may seem the reverse of quantum cloning, it is not so. Despite the distinction between these two operations there may be some link between the optimal fidelities of approximate deleting and cloning. Nevertheless, nature seems to put another limitation on quantum information imposed by the linearity of quantum mechanics. $\square$

Received 6 December 1999; accepted 7 January 2000.

1. Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
2. Dieks, D. Communication by EPR devices. *Phys. Lett. A* **92**, 271–272 (1982).
3. Bužek, V. & Hillery, M. H. Quantum copying: Beyond the no-cloning theorem. *Phys. Rev. A* **54**, 1844–1852 (1996).
4. Bužek, V., Braunstein, S. L., Hillery, M. H. & Bruß, D. Quantum copying: A network. *Phys. Rev. A* **56**, 3446–3452 (1997).
5. Duan, L. M. & Guo, G. C. A probabilistic cloning machine for replicating two non-orthogonal states. *Phys. Lett. A* **243**, 261–264 (1998).
6. Duan, L. M. & Guo, G. C. Probabilistic cloning and identification of linearly independent states. *Phys. Rev. Lett.* **80**, 4999–5002 (1998).
7. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191 (1961).
8. Yuen, H. P. Amplification of quantum states and noiseless photon amplifiers. *Phys. Lett. A* **113**, 405–407 (1986).
9. Pati, A. K. Quantum superposition of multiple clones and the novel cloning machine. *Phys. Rev. Lett.* **83**, 2849–2852 (1999).
10. Bennett, C. H. & Brassard, G. The dawn of a new era for quantum cryptography: the experimental prototype is working! *Special Interest Group on Automata and Computability Theory News* **20**, 78–82 (1989).
11. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
12. Boschi, D. *et al.* Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
13. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
14. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
15. Deutsch, D. Quantum-theory, the Church–Turing principle and the universal computer. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).
16. Gisin, N. & Massar, S. Optimal quantum cloning machines. *Phys. Rev. Lett.* **79**, 2153–2156 (1997).

Acknowledgements

We thank C. H. Bennett, S. Popescu, S. Bose, L. M. Duan and N. J. Cerf for useful comments.

Correspondence and requests for materials should be addressed to A.K.P. (e-mail: akpati@sees.bangor.ac.uk).

Molecular control over Au/GaAs diodes

Ayelet Vilan*, Abraham Shanzer† & David Cahen*

* Department of Materials and Interfaces, † Department of Organic Chemistry, Weizmann Institute of Science, Rehovoth 76100, Israel

The use of molecules to control electron transport is an interesting possibility, not least because of the anticipated role of molecules in future electronic devices¹. But physical implementations using discrete molecules are neither conceptually^{2,3} simple nor technically straightforward (difficulties arise in connecting the molecules to the macroscopic environment). But the use of molecules in electronic devices is not limited to single molecules, molecular wires or bulk material. Here we demonstrate that molecules can control the electrical characteristics of conventional metal–semiconductor junctions, apparently without the need for electrons to be transferred onto and through the molecules. We modify diodes by adsorbing small molecules onto single crystals of n-type GaAs semiconductor. Gold contacts were deposited onto the modified surface, using a ‘soft’ method to avoid damaging the molecules⁴. By using a series of multifunctional molecules whose dipole is varied systematically, we produce diodes with an effective barrier height that is tuned by the molecule’s dipole moment. These barrier heights correlate well with the change in work function of the GaAs surface after molecular modification. This behaviour is consistent with that of unmodified metal–semiconductor diodes, in which the barrier height can depend on the metal’s work function.

The lack of clear understanding of electron transport involving non-conjugated molecules⁵ contrasts sharply with the rather detailed understanding of transport in solids with extended bonding (non-molecular ones), also on a mesoscopic scale. Most work in this area is concerned with electron transport through molecules^{6,7}, and several mechanisms have been proposed for such transport⁸.

The experiments reported here are based on our ability to use series of molecules for systematic modification of the surface electronic properties of semiconductors and metals^{9–14}. In terms of actual electron transport, we used such molecules to explore molecular control over CuInSe₂/CdS (ref. 15) and CdTe/Au (ref. 16) diode characteristics. Characteristics of organic LEDs were changed by adsorption of nitrobenzoic acid and a carboxylated Ru dye¹⁷, or of thiols with two different dipoles¹⁸. Krueger *et al.* showed¹⁹ that the current–voltage (*I*–*V*) characteristics of diodes (made by spin-coating an organic hole conductor on thin polycrystalline TiO₂ films), onto which a series of benzoic acids was adsorbed, change with the dipole moments of the benzoic acid used.

A significant problem in all such studies is deposition of the second contact on the molecularly modified surface^{20,21}. Techniques used to make ordinary types of diodes, such as sputtering and vacuum evaporation, can (and often will) damage the molecules and/or change the surface in an uncontrollable fashion (see ref. 4 for a review of the various methods tried for contact deposition on organic molecular monolayers). The encouraging preliminary results obtained for less common types of diodes, made with milder methods (for example, electrodeposition of CdS onto molecularly modified CuInSe₂; ref. 15) led us to search for ways to modify a more usual type of diode with molecules. Thus, we looked for single-crystal semiconductor–metal junctions to which the metal contact can be applied as gently as possible.

We used the series of tartaric acid derivatives, whose adsorption characteristics to GaAs(100) we characterized earlier. The molecules bind strongly to GaAs (pK = 4–6) via a dicarboxylic acid group¹³. A substituted benzene group is attached to the binding group, which

allows systematic modification of the molecule’s dipole (Fig. 1, inset). The molecules self-assemble onto the surface into roughly one monolayer by overnight adsorption in a 2.5 mM acetonitrile solution onto etched¹⁰ n-GaAs(100)^{13,14}. Indication of surface coverage comes from saturation of both spectroscopic (Fourier transform infrared) and electrical (surface resistance) signals. Fitting to a Langmuir isotherm shows a continuous fit over the complete range of concentration and time¹³. Hence, multilayer adsorption is very unlikely. From contact potential difference (CPD) measurements (see below), the coverage is found to be at least 2/3 of a monolayer, in accordance with results for related molecules, on Au (ref. 9). Wafers were purchased from AXT Co., and Si-doped to ~10¹⁸ cm⁻³. Use of relatively highly doped GaAs assured reproducible preparation of good ohmic back contacts. The strong chemisorption of the molecules stabilizes the GaAs surface that is otherwise notoriously difficult to control. Hence, experiments with molecularly modified surfaces can be done at ambient conditions¹⁰.

Gold contacts were deposited on the modified surface in a very ‘soft’ manner. For this we adapted the ‘lift-off, float-on’ technique (LOFO)²⁰—used, for example, to prepare samples for transmission electron microscopy and to deposit epitaxial GaAs films on foreign substrates²²—for use with organic solvents instead of water. Au dots, 60 nm thick and 0.5 mm in diameter, were evaporated onto clean glass slides, from which they were allowed to peel by dipping the slide at an angle in 5 vol.% solution of HF in acetonitrile. The solvent was then changed to pure acetonitrile, which contains the modified GaAs crystal; 0.5–1 vol.% water is added to increase the surface tension, to allow the Au leaves to float. The GaAs is then lifted with the Au on top, and the samples are dried at 60–70 °C in a vacuum oven. The morphology of the contacts (at micrometre resolution) depends strongly on the floating solvent; in general, the more hydrophobic the chemisorbed molecules, the smoother the contacts. Ohmic back contact to the GaAs was made by rubbing InGa eutectic onto it. Current–voltage (*I*–*V*) characteristics were measured using a W needle, connected to a micromanipulator to

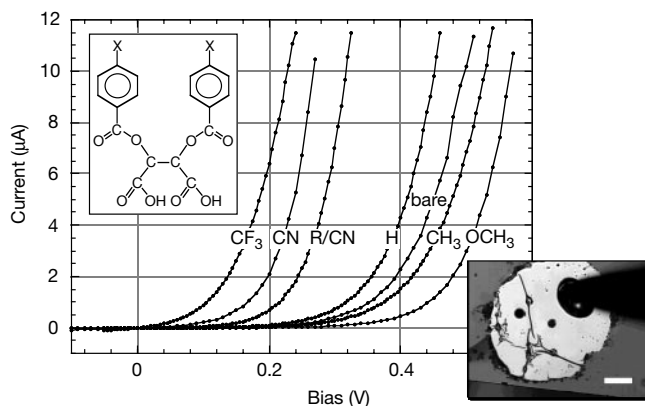


Figure 1 Characteristics of Au/molecule-on-n-GaAs junctions. Main figure, the different current–voltage (*I*–*V*) curves refer to junctions that are identical except for the substituent on the benzene ring (shown next to each curve) of the dicarboxylic acid derivatives (see inset) that were chemisorbed onto the n-GaAs surface. Each curve is typical for several measurements. No breakdown of devices was observed up to 40 mA cm⁻² and 1 V forward bias; at current densities >5 mA cm⁻², *n* increased to 5–7. The ‘bare’ curve is for a reference sample that was immersed in pure solvent, without adsorbed molecules. Because it was exposed to air after LOFO contacting, this is an oxidized surface¹⁰ without molecules. Top left inset, general chemical formulae for the molecules, chemisorbed on GaAs. The substituent, X, appears explicitly next to each curve in the main panel; R/CN is a unique ligand with asymmetric arms, one of which is a CN-substituted benzene ring, while the other is a C₁₅ alkyl chain instead of a benzene ring (compare ref. 12). Bottom right inset, optical micrograph (scale bar, 100 μm) of Au pad²⁰, floated on molecularly modified (X = CF₃) GaAs. The needle and the InGa drop (see text) can be seen on the pad (upper right corner of insert).

contact the Au dot (an InGa drop on the Au minimizes mechanical (pressure) damage to the film), and an HP 4155 semiconductor parameter analyser, in the voltage scan mode; these measurements were made mostly at ambient conditions.

I - V curves for a series of molecularly modified Au/n-GaAs junctions are shown in Fig. 1. They show qualitatively that the organic monolayer increases the current if the dipole is positive ($-\text{CF}_3$ and $-\text{CN}$ substituents), and decreases the current for negative dipoles ($-\text{CH}_3$ and $-\text{OCH}_3$), compared to the unmodified junction. Hence, a monolayer that is not providing perfect coverage of the surface (that is, with a high probability of pinholes)—made up of molecules, themselves expected to be poor electrical conductors, and so thin that electron tunnelling through them is relatively easy—tunes the diode current.

To understand this, we note that the junction's barrier height (ϕ_b), is influenced by the interface dipole (μ) at the GaAs–Au contact. In an ideal Schottky junction, the barrier height is the difference between the metal work function ($q\phi_{\text{metal}}$, with q the absolute value of the electron charge in coulombs) and the semiconductor electron affinity (χ_{sc}). The work function is a characteristic of the surface of the material, which changes with changes in the surface dipole^{23,24}. We can view χ_{sc} as the sum of a constant (χ_{sc}^0) and a dipole ($q\phi_{\text{dipole}}$) contribution. Therefore, the net barrier height is:

$$q\phi_b = q\phi_{\text{metal}} - (\chi_{\text{sc}}^0 + q\phi_{\text{dipole}}) \quad (1)$$

Here ϕ_{dipole} is the potential step due to interface dipoles²²:

$$\phi_{\text{dipole}} = \frac{N\mu \cos \theta}{\epsilon\epsilon_0} \quad (2)$$

where μ is the dipole moment (in units of C m), N is the surface density of dipoles (in m^{-2}), tilted at an (average) angle θ , ϵ is the layer's dielectric constant, and ϵ_0 is the dielectric permittivity of vacuum.

A dipole at the interface between a semiconductor and a conductor will induce or change the space charge, and thus the electric potential difference between the conductor and the bulk of the semiconductor^{23–25} (that is, the junction's barrier height, compare Fig. 8.9 in ref. 26). This is in marked contrast to what happens on a free surface, where only net surface charge, associated with surface states, can bring about such potential difference between the semiconductor surface and bulk. In the case of an interface, the effect will be there even in the absence of interface states. Such effects are expected with unintentional surface dipoles (for example, those due to surface reconstruction or to chemical metal–semiconductor interaction) and with intentionally applied ones. Experimentally, the effect of interface dipoles has been found, for example, in heterojunctions by variation of the doping at the interface²⁷. Based on this, we can expect the molecules to have a similar effect. To verify this, we fitted the I - V characteristics to the Schottky thermionic emission formula, also often used as an empirical way to describe current transport across a barrier^{23,26}:

$$I = I_s \exp\left(\frac{qV}{nkT}\right) \left(1 - \exp\left(-\frac{qV}{kT}\right)\right) \quad (3)$$

Here V is the applied bias voltage (in V), I the measured current (in A), k and T are the temperature and Boltzmann's constant, respectively, and n is the so-called ideality factor which corrects for deviation from ideal thermionic emission behaviour. I_s is the saturation current given by:

$$I_s = \sigma A^* T^2 \exp\left(-\frac{q\phi_b}{kT}\right) \quad (4)$$

where ϕ_b is now the effective barrier height (in V); A^* is the Richardson constant (in $\text{A cm}^{-2} \text{K}^{-2}$) and σ is the contact area (in cm^2)²⁸.

We note that n can be extracted from the slope of $(\ln I) - V$ plots, and the ϕ_b can be calculated from the intercept of $\ln I$ at zero bias; for our data, n was in the range 1.3–2.3 which means that the transport mechanism is not solely one of thermionic emission^{23,26,28}. Advanced analyses could consider additional mechanisms for carrier transport across the molecularly modified interface: for example, via tunnelling (across the space charge, oxidized surface and molecular layers). Because metal–semiconductor junctions are majority carrier devices²⁹, in these Au/n-GaAs structures the carriers are expected to be primarily electrons.

The barrier height obtained here is an effective one, ignoring barrier lowering due to the GaAs surface layer, image forces²⁸, and

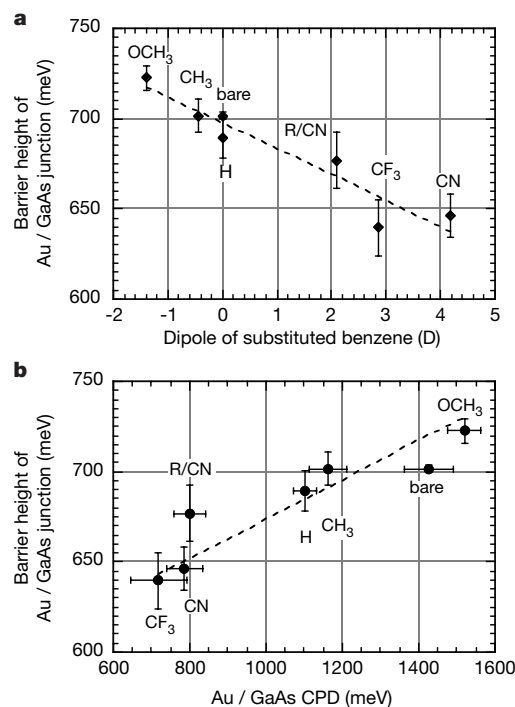


Figure 2 Dependence of the effective barrier height (ϕ_b) at the Au/derivatized n-GaAs interface on the following parameters. **a**, The dipole moment; **b**, the contact potential difference (CPD) between Au and derivatized n-GaAs surfaces. Values of the effective barrier height were extracted from I - V curves (compare Fig. 1), using equations (3) and (4). For each ligand, several samples were measured, on each of which two to four different contact pads were used, with about three measurements per pad. No significant changes were seen between 6–10 successive measurements. The error bars in the plot are the standard deviations of the measured data. We used only those curves that gave a good fit to the modified diode equation²⁸:

$$\ln\left[\frac{I}{1 - e^{-\frac{qV}{kT}}}\right] = \ln(\sigma A^* T^2) + \frac{q}{kT} \left(-\phi_b + \frac{V}{n}\right)$$

That is, we used only curves that met the requirement of continuity around zero bias, up to a negative bias of 0.1 V. On a semi-log plot (not shown), the curves are linear from -0.1 to about $+0.5$ V. The dipole moment refers to the substituted benzene part of the dicarboxylic acid^{9,12,32}. Note that the dipole of the substituted benzene moiety and the complete molecule are not identical, as the molecular dipole is the sum of those of the benzene moiety, the binding group, and the Ga–O bond. This issue is discussed in detail in refs 9 and 12. The direction of the dipole is taken as positive for the negative pole directed away from the semiconductor bulk. In **b**, the CPD was measured experimentally in an inert atmosphere using the Kelvin probe technique. The semiconductor's electron affinity, rather than its work function, was measured by using saturation illumination, that is, under conditions that minimize free surface band bending. The calculated ϕ_b value is smaller than expected without an organic layer (bare, oxidized surface), and larger than expected with the C₁₅ alkyl chain (R/CN ligand). These deviations may be ascribed to corresponding changes in the thickness and permittivity of the insulating layer²⁸, something that is being investigated further.

inhomogeneity of the barrier (compare Ch. 3 in ref. 23). The values are obtained using the nominal contact area ($\sigma = 0.002 \text{ cm}^2$) and the published²⁹ value for the Richardson factor ($A^* = 4 \text{ A cm}^{-2} \text{ K}^{-2}$). Thus, in this first approximation we neglect any (possible) effect of the molecules on the pre-exponential factor in equation (4), and attribute all of the molecular effect to the effective barrier height.

Values of ϕ_b of the molecularly modified Au/GaAs junction, extracted from the data shown in Fig. 1, using equations (3) and (4), are shown in Fig. 2a, as a function of the dipole moment of the substituted benzene group. As predicted from equations (1) and (2), the more positive the dipole, the higher the work function of GaAs, and the smaller ϕ_b .

An alternative way to measure the molecular effect on surface electronic properties is determination of the contact potential difference (CPD) between the GaAs surface and Au. We did this using a so-called Kelvin probe^{9,18,24}, which does not require physical contact with the modified surface. GaAs sample preparation was essentially the same as that used for preparing junctions. Figure 2b shows a comparison between ϕ_b of the modified Au/GaAs junction (on the y axis) and the CPD between Au and the modified 'free' GaAs surface (on the x axis).

Only 10% of the changes in CPD are expressed in the junctions' effective barrier heights. The reason is that, as written, equation (1) holds only rarely. Most semiconductors show appreciable band bending at their free surface, due to intrinsic surface states. If the density of these states is very high then they, rather than the difference in work function with the contact material, will dictate the junction's ϕ_b . This is the so-called Bardeen limit, in contrast to the Schottky–Mott limit expressed by equation (1) (refs 23, 28). In general, the less ionic the semiconductor, the less Schottky-like and the more Bardeen-like it behaves. Kurtin *et al.* defined an index of interface behaviour, S , which expresses the percentage of the work function difference that will be reflected in the barrier height³⁰. For GaAs, $S \approx 0.1$, which fits the 10% relative effect of the change in work function on ϕ_b found here.

We have shown that adsorbed molecular monolayers can tune the effective barrier height at the Au/GaAs interface, by their dipole effect. We note that such dipole effects will have to be considered when interpreting data from electron transport measurements, involving semiconductor/molecule interfaces³¹. Our results also suggest that we can tune the properties of electronic devices by way of molecular dipoles, rather than by actual electron transport through the molecules. Such an approach would allow the use of imperfect, practical molecular monolayers—with relatively easy connection to the outside world—which might fit the forecast of a future molecular central processing unit operating by electrostatic interactions rather than electron currents³. □

Received 20 October 1999; accepted 14 February 2000.

- Aviram, A. & Ratner, M. (eds) Molecular electronics: science and technology. *Ann. NY Acad. Sci.* **852**, (1998).
- Heath, J. R., Kuekes, P. J., Snider, G. S. & Williams, R. S. A defect-tolerant computer architecture: Opportunities for nanotechnology. *Science* **280**, 1716–1721 (1998).
- Tour, J. M., Kozaki, M. & Seminario, J. M. Molecular scale electronics: A synthetic/computational approach to digital computing. *J. Am. Chem. Soc.* **120**, 8486–8493 (1999).
- Haag, R., Rampi, M. A., Holmlin, R. E. & Whitesides, G. M. Electrical breakdown of aliphatic and aromatic self-assembled monolayers used as nanometer-thick organic dielectrics. *J. Am. Chem. Soc.* **121**, 7895–7906 (1999).
- Wilson, E. K. DNA conductance convergence? *Chem. Eng. News* **77**(36), 43–48 (1999).
- Bumm, L. A., Arnold, J. J., Dunbar, T. D., Allara, D. L. & Weiss, P. S. Electron transfer through organic molecules. *J. Phys. Chem. B* **103**, 8122–8127 (1999).
- Slowinsky, K., Fong, H. K. Y. & Majda, M. Mercury–mercury tunneling junctions. 1. Electron tunneling across symmetric and asymmetric alkanethiolate bilayers. *J. Am. Chem. Soc.* **121**, 7257–7261 (1999).
- Ratner, M. A. *et al.* Molecular wires: charge transport, mechanisms, and control. *Ann. NY Acad. Sci.* **852**, 22–37 (1998).
- Bruening, M. *et al.* Simultaneous control of surface potential and wetting of solids with chemisorbed multifunctional ligands. *J. Am. Chem. Soc.* **119**, 5720–5728 (1997).
- Bastide, S. *et al.* Controlling the work function of GaAs by chemisorption of benzoic acid derivatives. *J. Phys. Chem.* **101**, 2678–2684 (1997).
- Cohen, R., Zenou, N., Cahen, D. & Yitzchaik, S. Molecular electronic tuning of Si surfaces. *Chem. Phys. Lett.* **279**, 270–274 (1997).

- Cohen, R. *et al.* Controlling electronic properties of CdTe by adsorption of dicarboxylic acid derivatives: Relating molecular parameters to band bending and electron affinity changes. *Adv. Mater.* **9**, 746–749 (1997).
- Vilan, A. *et al.* Real time electronic monitoring of adsorption kinetics; evidence for 2-site adsorption mechanism of dicarboxylic acids on GaAs (100). *J. Phys. Chem.* **102**, 3307–3309 (1998).
- Cohen, R. *et al.* Controlling semiconductor surface electronic properties by dicarboxylic acids. *J. Amer. Chem. Soc.* **121**, 10545–10551 (1999).
- Gal, D. *et al.* Engineering the interface energetics of solar cells by grafting molecular properties onto semiconductors. *Proc. Ind. Acad. Sci. Chem. Sci.* **109**, 487–496 (1997).
- Moons, E. *Linking the interfacial chemistry and physics of CuInSe₂- and CdTe-based photovoltaic cells and diodes* PhD thesis Feinberg Grad. School, Weizmann Institute of Science, Rehovot (1995).
- Zuppiroli, L. *et al.* self-assembled monolayers as interfaces for organic opto-electronic devices. *Eur. Phys. J. B* **11**, 505–512 (1999).
- Campbell, L. H. *et al.* Controlling charge injection in organic electronic devices using self-assembled monolayers. *Appl. Phys. Lett.* **71**, 3528–3530 (1997).
- Krueger, J., Bach, U. & Grätzel, M. Modification of TiO₂ heterojunctions with benzoic acid derivatives in hybrid molecular solid-state devices. *Adv. Mater.* **12**, 447–451 (2000).
- Moons, E. *et al.* Electron transport in hybrid organic/inorganic structures. *Synth. Met.* **76**, 245–248 (1996).
- Martin, A., Sambles, J. & Ashwell, G. Molecular rectifier. *Phys. Rev. Lett.* **70**, 218–221 (1993).
- Yablonovitch, E., Hwang, D. M., Gmitter, T. J., Florez, L. T. & Harbison, J. P. Van der Waals bonding of GaAs epitaxial liftoff films onto arbitrary substrates. *Appl. Phys. Lett.* **56**, 2419–2421 (1990).
- Brillson, L. *J. Contacts to Semiconductors; Fundamentals and Technology* (Noyes, Park Ridge, New Jersey, 1993).
- Kronik, L. & Shapira, Y. Surface photovoltage techniques: theory, experiments and applications. *Surf. Sci. Rep.* **37**, 1–204 (1999).
- van Ruyven, L. J. The position of the Fermi level at a heterojunction interface. *Phys. Status Solidi* **5**, K109–K111 (1964).
- Boer, K. W. *Survey of Semiconductor Physics Vol. II, Barriers, Junctions, Surfaces, and Devices* (Van Nostrand Reinhold, New York, 1992).
- Capasso, F., Cho, A. Y., Mohammed, K. & Foy, P. W. Doping interface dipoles: Tunable heterojunction barrier heights and band-edge discontinuities by molecular beam epitaxy. *Appl. Phys. Lett.* **46**, 664–666 (1985).
- Rhoderick, E. A. & Williams, R. H. *Metal-semiconductor Contacts* (Clarendon, Oxford, 1988).
- Sze, S. M. *Physics of Semiconductor Devices* (Wiley, New York, 1981).
- Kurtin, S., McGill, T. C. & Mead, C. A. Fundamental transition in the electronic nature of solids. *Phys. Rev. Lett.* **22**, 1433–1436 (1970).
- Yaliraki, S., Roitberg, A., Gonzalez, C., Mujica, V. & Ratner, M. The injecting energy at molecule/metal interfaces: Implications for conductance of molecular junctions from an ab initio molecular description. *J. Chem. Phys.* **111**, 6997–7002 (1999).
- Lide, D. R. (ed.) *Handbook of Chemistry and Physics* 73rd edn, 9–50 (CRC Press, Boca Raton, Florida, 1992).

Acknowledgements

We thank R. Fleishman for help with early experiments, L. Kronik for discussions, and J. Krüger, U. Bach and M. Grätzel (EPFL) for sharing their unpublished results with us.

Correspondence and requests for materials should be addressed to D.C. (e-mail: david.cahen@weizmann.ac.il).

Sintering dense nanocrystalline ceramics without final-stage grain growth

I.-Wei Chen & X.-H. Wang

Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6272, USA

Sintering is the process whereby interparticle pores in a granular material are eliminated by atomic diffusion driven by capillary forces. It is the preferred manufacturing method for industrial ceramics. The observation of Burke and Coble^{1,2} that certain crystalline granular solids could gain full density and translucency by solid-state sintering was an important milestone for modern technical ceramics. But these final-stage sintering processes are always accompanied by rapid grain growth^{3–6}, because the capillary driving forces for sintering (involving surfaces) and grain growth (involving grain boundaries) are comparable in magnitude, both being proportional to the reciprocal grain size.

This has greatly hampered efforts to produce dense materials with nanometre-scale structure (grain size less than 100 nm)^{4,8}, leading many researchers to resort to the 'brute force' approach of high-pressure consolidation at elevated temperatures⁷⁻⁹. Here we show that fully dense cubic Y₂O₃ (melting point, 2,439 °C) with a grain size of 60 nm can be prepared by a simple two-step sintering method, at temperatures of about 1,000 °C without applied pressure. The suppression of the final-stage grain growth is achieved by exploiting the difference in kinetics between grain-boundary diffusion and grain-boundary migration. Such a process should facilitate the cost-effective preparation of other nanocrystalline materials for practical applications.

For Y₂O₃, the normal sintering practice is to heat the powder compact at a certain rate, holding it at the highest temperature until the maximum density is reached. As shown in Fig. 1, the grain size of Y₂O₃ samples sintered in this way increases continuously with the density. Depending on the composition, the final grain size varies between 200 and 600 nm, with most of the variation occurring when the density exceeds 80%. This variation is dopant-dependent: MgO enhances grain growth while Nb₂O₅ suppresses it, which is consistent with the reported grain-boundary mobilities for these compositions¹⁰. These data are in line with common sintering experience: accelerated grain growth accompanies final-stage sintering.

The sintering method that we report here uses two steps in the heating schedule. The sample is first heated to a higher temperature to achieve an intermediate density, then cooled down and held at a lower temperature until it is fully dense. For pure Y₂O₃ initially heated to 1,310 °C, full density was achieved after holding at 1,150 °C for 20 h, during which there was no grain growth (Fig. 2). As another example, with the first step reaching 1,250 °C, full densification without grain growth was achieved after holding at 1,150 °C for 20 h. These results are entirely different from those of Fig. 1: sintering has not resulted in grain growth.

To succeed in two-step sintering, a sufficiently high starting density should be obtained during the first step. When the density is above 70%, porosimetry data¹¹ have shown that all pores in Y₂O₃ become subcritical and unstable against shrinkage¹¹⁻¹³ (which

occurs by capillary action). These pores can be filled as long as grain-boundary diffusion allows it, even if the particle network is frozen as it clearly is in the second step. In our study, we found that densities higher than 75% were adequate for subsequent second-step sintering.

The lack of grain growth in second-step sintering has important implications for kinetics. Grain coarsening creates a powerful dynamic that constantly refreshes the microstructure. Statistically, only one-eighth of all grains survive every time the size of the grains doubles. This evolution can be a source of enhanced kinetics. Even without grain growth, enhanced kinetics has also been suspected in cases when microstructure evolution is otherwise robust; for example, in fine-grain superplasticity¹⁴⁻¹⁶. Because second-step sintering proceeds in a 'frozen' microstructure, it should have slower kinetics. Yet the slower kinetics is sufficient for reaching full density, while providing the benefit of suppressing grain growth.

We quantified the diffusion kinetics in the frozen microstructure by measuring the densification rates in the second step, and comparing them with the prediction based on Herring's dimensional argument¹⁷ for normalized densification rate ($d\rho/\rho dt$, where ρ is relative density and t is time):

$$d\rho/\rho dt = F(\rho)(\gamma\Omega/GkT)(\delta D/G^3) \quad (1)$$

Here γ is surface energy, Ω is atomic volume, G is grain size, δ is grain-boundary thickness, and D is grain-boundary diffusivity. In the above, $\gamma\Omega/GkT$ may be viewed as the normalized driving force, and $\delta D/G^3$ is the standard kinetic factor that enters the strain-rate equation for grain-boundary processes such as sintering, diffusional creep and superplasticity. The remaining dimensionless prefactor on the right-hand side, F , is independent of the grain size as such, but could depend on other aspects of the microstructure such as density and pore distribution. In our experiments, the grain network is frozen. Therefore, the only microstructure parameter that enters F during the second-step sintering should be density itself. To verify this, we plot in Fig. 3 normalized sintering rates against density for different runs of second-step sintering; we note that the data fall onto parallel curves in the semi-log plot. This implies that the only difference between different runs is in the timescale, and

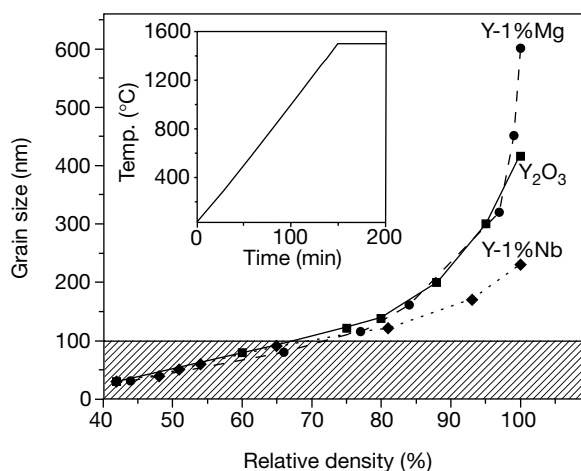


Figure 1 Increasing grain size of Y₂O₃ with density in normal sintering. (Heating schedule shown in inset). Even with fine starting powders (30 nm), the final grain size of dense ceramics is well over 200 nm regardless of whether dopant was used or not. Shaded area indicates the grain size regime commonly defined as nanostructured materials. At lower densities, the mean grain (particle) size was estimated on the fracture surface. At higher densities, the grain size was obtained by multiplying by 1.56 the average linear intercept length of at least 500 grains.

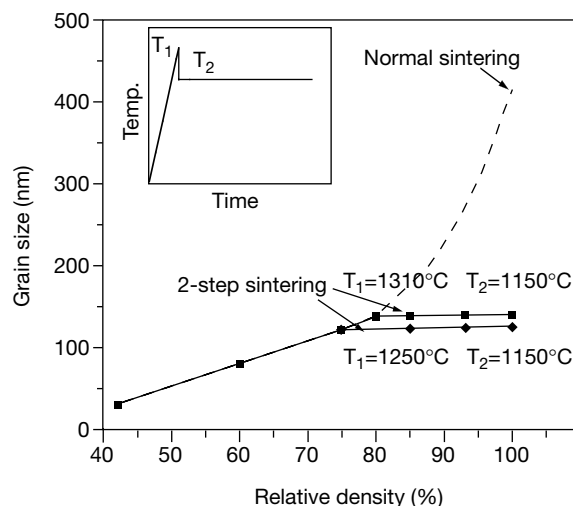


Figure 2 Grain size of Y₂O₃ in two-step sintering. (Heating schedule shown in inset.) Note that the grain size remains constant in the second sintering step, despite density improvement to 100%.

that $F(\rho)$ is a 'universal' function for all frozen microstructures. It also shows that normalized sintering rate and $F(\rho)$ decrease by a factor of 30 as the density increases from 85% to 99%; this means that sintering in a frozen microstructure is an exhaustion process that dramatically slows with density. In contrast, previous studies^{18,19} using normal sintering schedules suggested that $F(\rho)$ either increased or decreased slightly with density. The decreased kinetics can not be due to changes in diffusion mechanisms as the density increases; otherwise, the curves in Fig. 3 at different temperatures would not be parallel to each other as they reflect the corresponding changes in activation energy. (The normalized sintering rate data in Fig. 3—and other data that we obtained, but which are not shown—verified the strong grain-size dependence predicted by equation (1), suggesting grain boundary diffusion

control.) More likely, the decreased kinetics arises from a progressively longer diffusion distance during constant-grain-size sintering, because pores having a shorter diffusion distance are eliminated first and those having longer diffusion distances remain and progressively take control.

The feasibility of densification without grain growth relies on the suppression of grain-boundary migration while keeping grain-boundary diffusion active. The kinetic 'window' for which this is possible in second-step sintering of pure Y_2O_3 is delineated in Fig. 4, where the filled symbols indicate conditions where we obtained full densification without grain growth. An upper temperature, above which grain growth was observed, bounds this window. This suggests that grain-boundary migration may involve an activation process that has a higher activation energy than grain-boundary diffusion itself; thus, it is active at higher temperatures but is suppressed at lower temperatures. (The driving force for grain growth decreases with grain size, so the temperature required to activate the higher-energy process increases with grain size. This explains why the upper temperature in Fig. 4 shifts to a higher value at a larger grain size.) Recent studies of high-purity zinc have shown that grain-boundary migration can be severely hampered by the slow mobility of grain junctions at lower temperatures, the latter having a higher activation energy²⁰. It is possible that a similar process, in which grain junctions as well as grain boundary/pore junctions impede grain-boundary migration, may here explain the apparent suppression of grain growth at lower temperatures. Figure 4 also shows a lower-bound temperature, below which sintering is exhausted before full density is achieved. This suggests that grain boundary diffusion itself can be suppressed at lower temperatures. Interface kinetics in very fine grain polycrystals is sometimes limited due to difficulties in maintaining sources and sinks to accommodate point defects^{21–23}. This leads to a threshold energy or stress, of the order of $2\gamma/G$. For a grain size of 100 nm, this amounts to 20 MPa, which is rather substantial compared to capillary pressure and could be the cause for the suppression. This effect should diminish at larger grain sizes, allowing the kinetic window to extend to lower temperatures. Lastly, we found that Nb_2O_5 doping had the effect of shifting the kinetic window to a higher temperature, while MgO doping had the opposite effect. (See the dashed lines and the dotted lines in Fig. 4.) Therefore, in exploiting the difference in the kinetics of grain-boundary diffusion and grain-boundary migration to achieve densification without growth at lower temperatures, it is still advisable to utilize dopants to 'tune' the overall kinetics.

The finest grain size of fully dense ceramics examined so far lies around 60 nm (see Fig. 5), and is about 4–6 times the starting

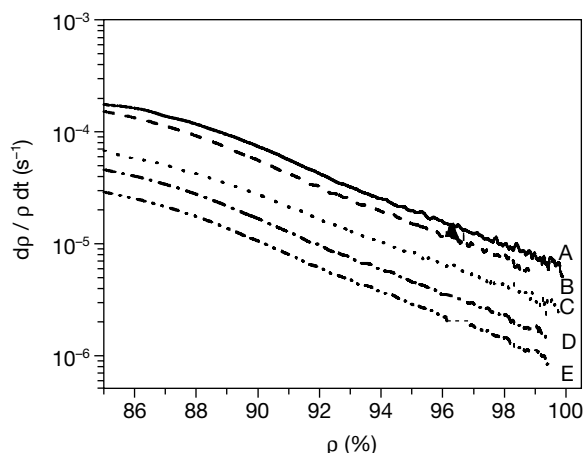


Figure 3 Normalized densification rate versus relative density during second-step sintering. (Heating schedule same as in Fig. 2 inset.) Trace A, Y_2O_3 : 1,250 °C then 1,150 °C for 20 h (final grain size, 122 nm). B, Y_2O_3 : 1,210 °C then 1,150 °C for 20 h (final grain size, 140 nm). C, Y_2O_3 (1% Nb): 1,450 °C then 1,200 °C for 40 h (final grain size, 118 nm). D, Y_2O_3 : 1,310 °C then 1,100 °C for 30 h (final grain size, 140 nm). E, Y_2O_3 (1% Mg): 1,250 °C then 1,050 °C for 30 h (final grain size, 116 nm). All curves are parallel to each other.

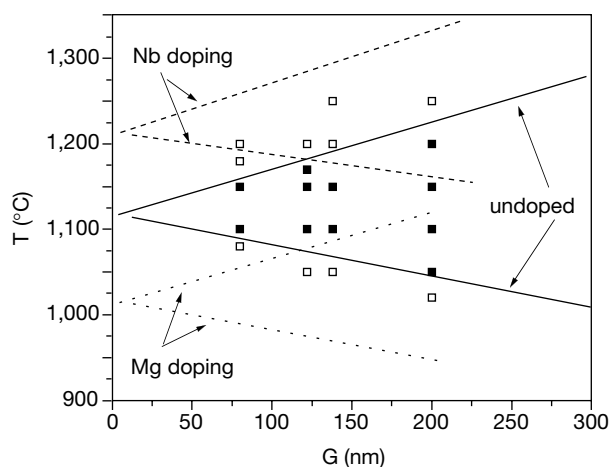


Figure 4 Kinetic window for reaching full density without grain growth. Solid symbols are temperature and grain-size conditions of successful second-step sintering runs for Y_2O_3 . Open symbols above the upper boundary are conditions that showed grain growth in second-step sintering, and open symbols below the lower boundary are ones that did not reach full density in second-step sintering. Similar boundaries are shown for Y_2O_3 doped with 1% Nb (dashed lines) and with 1% Mg (dotted lines).

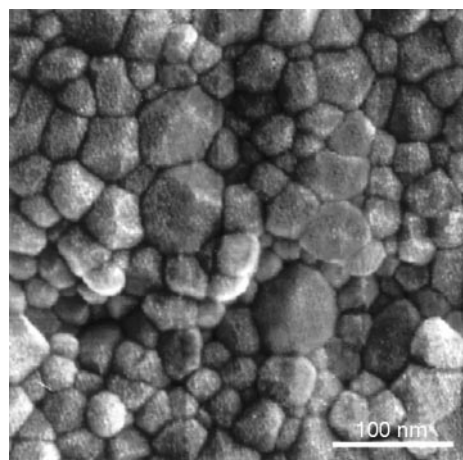


Figure 5 Microstructure of fully dense Y_2O_3 doped with 1% Mg. This material was sintered at 1,000 °C for 20 h, after reaching 76% relative density in an initial firing to 1,080 °C.

powder size (see Supplementary Information for details). This increase is entirely due to coarsening in the first-step sintering. As nanocrystalline powders in the size range 5–10 nm are becoming increasingly available^{24,25}, we believe that it should be possible to achieve dense, nanostructured materials of grain size 25–50 nm by exploiting the kinetic ‘window’ that separates grain-boundary diffusion from grain growth. We consider that the simplicity of this approach should also make it useful for detailed explorations of dense nanostructured materials, in order to take advantage of their grain-size-dependent physical properties. □

Methods

We used a precipitation technique to first obtain $Y(OH)_3$, which was then calcined at various temperatures from 600 to 800 °C to obtain final Y_2O_3 powders with a size from 10 to 100 nm. These powders were sifted and pressed into pellets. When doping was desired, Mg and Nb salts were introduced to aqueous slurries of Y_2O_3 powders; the slurries were gelled by adjusting their pH, then dried and recalcined. The pellets were sintered in a dilatometer or in a box furnace in air. The final density was determined using Archimedes’ method and quantitative microscopy.

Received 31 August 1999; accepted 17 January 2000.

- Coble, R. L. Sintering crystalline solids. *J. Appl. Phys.* **32**, 787–799 (1961).
- Coble, R. L. & Burke, J. E. in *Progress in Ceramic Science* Vol. 3 (ed. Burke, J. E.) 197–251 (New York, 1963).
- Brook, R. J. in *Treatise on Materials Science and Technology* Vol. 9 (ed. Wang, F. F. Y.) 331–363 (Academic, New York, 1976).
- Coble, R. L. & Cannon, R. M. in *Processing of Crystalline Ceramics* (ed. Palmour, H. III) 151–168 (Materials Science Research Vol. 11, Plenum, New York, 1978).
- Coble, R. L., Song, H., Brook, R. J., Handwerker, C. A. & Dynys, J. M. in *Structure and Properties of MgO and Al_2O_3 Ceramics* (ed. Kingery, W. D.) 839–852 (Vol. 10, Advances in Ceramics, American Ceramic Society, Westerville, OH, 1984).
- Cameron, C. P. & Raj, R. Grain growth transition during sintering of colloiddally prepared alumina powder compacts. *J. Am. Ceram. Soc.* **71**, 1031–1035 (1988).
- Gleiter, H. Nanocrystalline materials. *Prog. Mater. Sci.* **33**, 223–315 (1989).
- Mayo, M. J. Processing of nanocrystalline ceramics from ultrafine particles. *Int. Mater. Rev.* **41**, 85–115 (1996).
- Liao, S.-C., Chen, Y.-J., Kear, B. H. & Mayo, W. E. High pressure/low temperature sintering of nanocrystalline alumina. *Nanostruct. Mater.* **10**, 1063–1079 (1998).
- Chen, P.-L. & Chen, I.-W. Grain boundary mobility in Y_2O_3 : defect mechanism and dopant effects. *J. Am. Ceram. Soc.* **79**, 1801–1809 (1996).
- Chen, P.-L. & Chen, I.-W. Sintering of fine oxide powder: I, microstructural evolution. *J. Am. Ceram. Soc.* **79**, 3129–3141 (1996).
- Chen, P.-L. & Chen, I.-W. Sintering of fine oxide powders: II, sintering mechanisms. *J. Am. Ceram. Soc.* **80**, 637–645 (1997).
- Kingery, W. D. & Francois, B. in *Sintering and Related Phenomena* (eds Kuczynski, G. C., Hooton, N. S. & Gibbon, C. F.) 471–496 (Gordon & Breach, New York, 1967).
- McFadden, S. X., Mishra, R. S., Valiev, R. Z., Zhilyaev, A. P. & Mukherjee, A. K. Low-temperature superplasticity in nanostructured nickel and metal alloys. *Nature* **396**, 684–686 (1999).
- Wakai, F. et al. A superplastic covalent crystal composite. *Nature* **344**, 421–423 (1990).
- Chen, I.-W. & Xue, L.-A. Development of superplastic structure ceramics. *J. Am. Ceram. Soc.* **73**, 2585–2609 (1990).
- Herring, C. Effect of change of scale on sintering phenomena. *J. Appl. Phys.* **21**, 301–303 (1950).
- Zhao, J. & Harmer, M. P. Sintering of ultra-high purity alumina doped simultaneously with MgO and FeO . *J. Am. Ceram. Soc.* **70**, 860–866 (1987).
- Hansen, J. D., Rusin, R. P., Teng, M.-H. & Johnson, D.-L. Combined stage sintering model. *J. Am. Ceram. Soc.* **75**, 1129–1135 (1992).
- Czubayko, L., Sursava, V. G., Gottstein, G. & Shvindlerman, L. S. Influence of triple junctions on grain boundary motion. *Acta Mater.* **46**, 5863–5871 (1998).
- Land, T. A., Martin, T. L., Potapenko, S., Palmore, G. T. & De Yoreo, J. J. Recovery of surfaces from impurity poisoning during crystal growth. *Nature* **399**, 442–445 (1999).
- Cannon, R. M., Rhodes, W. H. & Heuer, A. H. Plastic deformation of fine-grained alumina: I. interface-controlled diffusional creep. *J. Am. Ceram. Soc.* **63**, 48–53 (1980).
- Arzt, E., Ashby, M. F. & Verrall, R. A. Interface-controlled diffusional creep. *Acta Metall.* **31**, 1977–1989 (1983).
- Kumar, K. P. et al. Densification of nanostructured titania assisted by a phase transformation. *Nature* **358**, 48–51 (1992).
- Klein, S., Winterer, M. & Hahn, H. Reduced pressure chemical vapor synthesis of nano-crystalline silicon carbide powders. *Adv. Mater.* **4**, 143–149 (1998).

Supplementary information is available on Nature’s World-Wide Web site or as paper copy from the London editorial office of *Nature*.

Acknowledgements

This work was supported by the US Department of Energy, Office of Basic Energy Sciences, using facilities supported by the NSF MRSEC at the University of Pennsylvania.

Correspondence and requests for materials should be addressed to I.-W.C. (e-mail: iweichen@seas.upenn.edu).

The influence of Antarctic sea ice on glacial–interglacial CO_2 variations

Britton B. Stephens*† & Ralph F. Keeling*

* Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093–0244, USA

Ice-core measurements indicate that atmospheric CO_2 concentrations during glacial periods were consistently about 80 parts per million lower than during interglacial periods¹. Previous explanations for this observation^{2–9} have typically had difficulty accounting for either the estimated glacial O_2 concentrations in the deep sea, $^{13}C/^{12}C$ ratios in Antarctic surface waters, or the depth of calcite saturation; also lacking is an explanation for the strong link between atmospheric CO_2 and Antarctic air temperature¹. There is growing evidence that the amount of deep water upwelling at low latitudes is significantly overestimated in most ocean general circulation models^{10,11} and simpler box models previously used to investigate this problem. Here we use a box model with deep-water upwelling confined to south of 55 °S to investigate the glacial–interglacial linkages between Antarctic air temperature and atmospheric CO_2 variations. We suggest that low glacial atmospheric CO_2 levels might result from reduced deep-water ventilation associated with either year-round Antarctic sea-ice coverage, or wintertime coverage combined with ice-induced stratification during the summer. The model presented here reproduces 67 parts per million of the observed glacial–interglacial CO_2 difference, as a result of reduced air–sea gas exchange in the Antarctic region, and is generally consistent with the additional observational constraints.

Most theories for low glacial atmospheric CO_2 concentrations rely on one of two mechanisms: an increase in the strength of the biological pump as a result of changes in the supply or utilization of nutrients or light^{2–6}, or an increase in the ocean’s alkalinity via coral reef dissolution⁷ or carbonate sediment interactions^{8,9}. Models that invoke biological pump changes are generally inconsistent with the magnitude and/or direction of glacial–interglacial changes in $^{13}C/^{12}C$ ratios and nutrients^{12,13}, and with the lack of widespread glacial deep-water anoxia. On the other hand, models invoking alkalinity changes predict large increases in the depth of the glacial lysocline which are not observed.

Our model is not based on productivity or alkalinity increases, but rather on changes in the rate of air–sea gas exchange as a result of increased sea-ice cover at high southern latitudes. By significantly limiting the sea-to-air CO_2 flux in the primary region for deep-water ventilation, expanded Antarctic sea ice during glacial times may trap relatively more carbon in the deep ocean, thereby reducing atmospheric CO_2 concentrations. This mechanism would increase the effectiveness of the biological pump at storing CO_2 in the deep ocean without actually increasing productivity, and thus would not necessarily require changes in nutrient concentrations or utilization efficiencies. Because dissolved O_2 equilibrates with the atmosphere more rapidly than CO_2 , as a result of its lack of buffering chemistry, the limitation imposed by glacial sea ice would have a lesser effect on O_2 entering the deep ocean than on CO_2 leaving, and thus would not produce deep anoxia. Furthermore, increased Antarctic sea ice would not produce a large lysocline shift. Finally, the direct connections between Antarctic temperature and sea-ice cover provide a possible explanation for the observed synchrony between Antarctic temperature and atmospheric CO_2 concentrations during deglaciation⁶.

† Present address: Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado 80309–0216, USA.

Increases in Antarctic sea ice have been included in three-dimensional ocean model simulations, but their potential to significantly affect atmospheric CO₂ in these models may be limited by defects involving excessive low-latitude upwelling of deep water. Toggweiler and Samuels¹⁰ have shown from ¹⁴C comparisons, and Gnanadesikan and Toggweiler¹¹ from silica flux comparisons, that coarse-resolution ocean models all overestimate the amount of deep water upwelling across the main thermocline at low latitudes in the Indian and Pacific oceans. Previous box models used to explore palaeo-CO₂ controls^{2–4,14,15} also overestimate deep-water upwelling at low latitudes. These models have from 15 to 40 Sv (1 Sv = 10⁶ m³ s⁻¹) of upwelling directly from the deep to the warm surface box, which is much greater than the combined upper limit of ~7 Sv indicated by tracer and transect studies in the Pacific¹⁰ and Indian¹⁶ oceans.

To investigate the influence of Antarctic sea ice in the absence of low-latitude deep-water upwelling, we employ the biogeochemical ocean model shown in Fig. 1. The fundamental difference between this and other box models of the ocean carbon cycle is that deep waters only return to the surface south of the Antarctic Polar Front (APF), consistent with the “reconfigured conveyor” of Toggweiler and Samuels¹⁰. We divide the upper ocean into a warm, low-nutrient surface (S) box and a cooler, higher-nutrient thermocline (T) box. We also separate the upwelling that feeds the Antarctic Bottom Water (AABW) formation (B) box from that which flows north in

the Antarctic surface (A) box to allow for higher nutrient concentrations in AABW¹⁷. We selected these divisions to optimize the model’s representation of the modern distribution of preformed nutrients and export production. However, combining either the B and A boxes or the S and T boxes does not change our results significantly.

Deep waters that upwell along the Antarctic Divergence, mix with surface waters, and then flow north are generally thought to sink again at the APF as part of the low-salinity Antarctic Intermediate Water (AAIW)¹⁸. Although there is considerable uncertainty^{19,20}, it appears that some of this recent deep water remains near the surface north of the APF where it convectively mixes with surface waters of low-latitude origin before sinking. To investigate model sensitivities to uncertainties in the fate of northward-flowing Antarctic surface waters, we allow for surface exposure of an adjustable fraction (F_a) of these waters in the subantarctic (SA) box.

After prescribing temperatures, salinities, water transports, and surface nutrients, and parametrizing the effects of biological production, air–sea gas exchange, and carbonate sediment interaction, we integrate the model to calculate the steady-state distribution of total CO₂ (ΣCO_2), the ¹³C/¹²C ratio ($\delta^{13}\text{C}$), total alkalinity (TA), dissolved O₂, and the atmospheric CO₂ concentration. Figure 1 depicts a solution to this model using input parameters representing the modern pre-industrial state. To assess the sensitivity of our model to variations in Antarctic ice cover, we have calculated

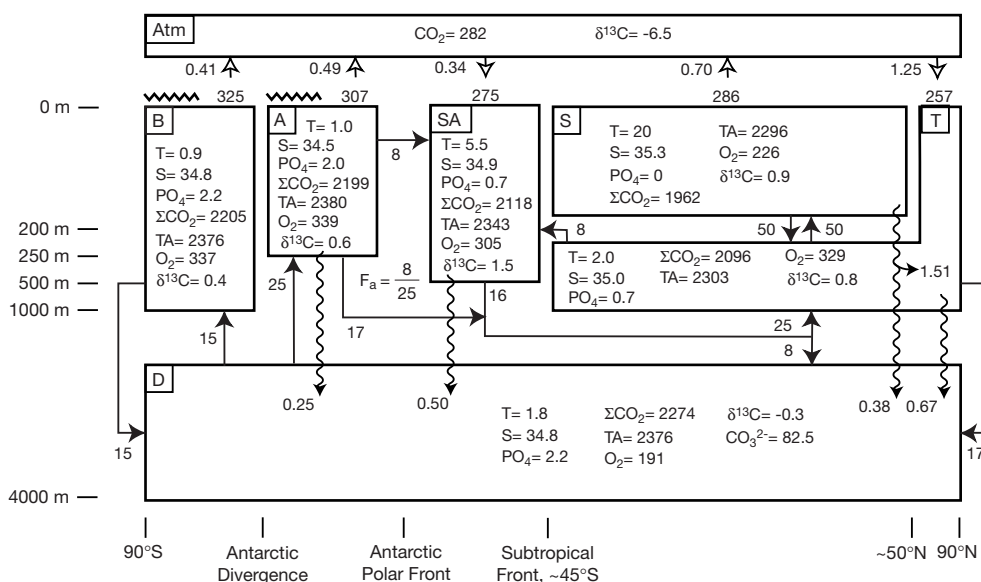


Figure 1 Solution from our atmosphere–ocean model for a best-guess modern-preindustrial state. In this state, $F_a = 0.3$, where F_a is the fraction of northward-flowing Antarctic surface waters exposed to the surface in the subantarctic (SA) box. This model consists of an atmosphere and six ocean boxes: a main surface box (S) representing the upper 200 m of water between approximately 50°N and the Southern Subtropical Front (STF), a main thermocline box (T) representing waters between 200 and 1,000 m depth with a surface outcrop north of the main surface box, an Antarctic Bottom Water formation box (B) representing the upper 1,000 m of water south of the Antarctic Divergence (AD) (~65°S today), an Antarctic surface box (A) representing the upper 250 m of the Southern Ocean south of the APF (~55°S today) and north of the AD, a subantarctic box (SA) representing the upper 500 m of water north of the APF and south of the STF, and a deep box (D) representing the remainder of the world’s oceans. Solid straight arrows denote water fluxes and are labelled in Sv. Solid wavy arrows indicate sinking fluxes of organic material and hollow arrows indicate air–sea CO₂ fluxes, and both are labelled in Gt C yr⁻¹. Numbers above the surface boxes denote CO₂ partial pressures in μatm . Values for PO₄, ΣCO_2 , TA, dissolved O₂, and CO₃²⁻ are indicated in $\mu\text{mol kg}^{-1}$, $\delta^{13}\text{C}$ in per mil, temperature T in °C, salinity S in p.s.u., and atmospheric CO₂ in p.p.m. We assume that 50% of the modern-preindustrial ocean in boxes B and A is covered by sea-ice

(represented by jagged lines), leaving a combined area of $1.6 \times 10^{13} \text{ m}^2$ open for the ventilation of upwelled deep waters. We prescribe surface phosphate concentrations in the A, SA, S and T boxes, and the total ocean phosphate concentration, such that the model reproduces the observed deep preformed phosphate (PO₄³⁻) concentration of $1.4 \mu\text{mol kg}^{-1}$ and the PO₄³⁻ concentration of intermediate waters penetrating the deep ocean, also equal to $1.4 \mu\text{mol kg}^{-1}$ (ref. 30). The model consumes excess phosphate in these four surface boxes at revised Redfield proportions of –175 O₂:127 C:16 N:1 P. As the resulting organic matter sinks, 80% of the main surface flux is oxidized in the thermocline box and the remainder of all three fluxes in the deep box. We assume inorganic to organic carbon production ratios of 0, 1:20, 1:10 and 1:3 in the A, SA, S and T boxes respectively. The model remineralizes all sinking calcium carbonate in the deep box. We calculate fluxes across the air–sea interface assuming a perfectly mixed atmosphere and using CO₂ invasion rates of 0.15 (B, A, SA, T) and 0.05 (S) mol m⁻² yr⁻¹ μatm^{-1} , and O₂ gas transfer velocities of 45 (B, A, SA, T) and 15 (S) cm h⁻¹. We maintain a constant CO₃²⁻ concentration in the deep box by dissolving or precipitating an appropriate amount of CaCO₃ (ref. 8). We have adjusted the mixing between the thermocline and surface boxes to produce a reasonable partitioning of biological productivity between these boxes.

steady-state solutions using these parameters and varying the ice-free surface area south of the APF from $1.6 \times 10^{13} \text{ m}^2$ down to zero. Figure 2a shows the atmospheric CO_2 concentration for these solutions for four cases with different values of F_a .

The decrease in atmospheric CO_2 of 67 parts per million (p.p.m.) for the best-guess conditions shown in Fig. 2 is associated with a 92% decrease in the sea-to-air CO_2 flux south of the APF, and a 1.8% increase in the deep ΣCO_2 concentration, with no change in the nutrient distribution. The deep ΣCO_2 increase contributes to the dissolution of $2.5 \times 10^{16} \text{ mol}$ of CaCO_3 , which produces a 1.6% increase in whole ocean alkalinity and contributes 13 p.p.m. of the total atmospheric CO_2 effect. Although not enough to produce large changes in glacial lysocline depths, this alkalinity-ice relationship is consistent with observations of a global carbonate-sediment preservation spike during deglaciation²¹.

In contrast to atmospheric CO_2 , the modelled deep O_2 concentration is not sensitive to ice coverage except when the outcrop area becomes very small (Fig. 2b). The change in deep O_2 from a modern-preindustrial estimate of $191 \mu\text{mol kg}^{-1}$ to a possible glacial value of $123 \mu\text{mol kg}^{-1}$ (Fig. 2b) would not produce deep anoxia. Our model also supports earlier suggestions^{12,22} that sea ice is an important influence on glacial Antarctic surface $\delta^{13}\text{C}$ values. Fractionation effects associated with both net and gross air-sea CO_2 fluxes work to make the Antarctic surface box richer in ^{13}C . As these fluxes decrease with increasing ice cover, the Antarctic surface $\delta^{13}\text{C}$ value decreases by 0.7‰ (Fig. 2b). This decrease is similar to that inferred from measurements on planktonic foraminifera¹², and is in contrast to the increases in Antarctic surface $\delta^{13}\text{C}$ implied by models that invoke increases in high-latitude nutrient utilization or decreases in high-latitude vertical mixing.

Atmospheric $\delta^{13}\text{C}$ increases with increasing sea ice by 0.9‰ in the best-case model scenario. If we include an input of 500 gigatons of terrestrial carbon²³, and reduce surface temperatures by a maximum of 5°C (ref. 24) and less in boxes that are already near the freezing point—the atmospheric $\delta^{13}\text{C}$ value returns to within 0.1‰ of the modern-preindustrial value. Measurements^{15,25} indicate a sharp dip in atmospheric $\delta^{13}\text{C}$ of $\sim 0.5\text{‰}$ at the start of termination I, followed by a more gradual increase to a pre-industrial value several tenths of a per mil above that at the Last Glacial Maximum (LGM). Although our model does not reproduce this overall shift, it does suggest that the temporal behaviour of atmospheric $\delta^{13}\text{C}$ during deglaciation could be explained by an initial meltback of Antarctic sea ice followed by a more gradual growth in terrestrial biomass and increase in surface temperatures.

Although it is difficult to estimate the extent of glacial Antarctic sea ice, suggestions from sediment proxies of significantly increased coverage justify the investigation of its potential implications for atmospheric CO_2 . There is solid evidence from ice-rafted volcanic detritus²⁶, and support from fossil plankton assemblages²⁷, that wintertime Antarctic sea ice extended to or beyond the modern APF during the LGM. However, summertime LGM ice limits are not as well constrained. Early estimates of significantly increased coverage based on sediment types²⁶ have been countered by recent planktonic analyses that suggest summertime Antarctic sea ice at the LGM may not have been much more extensive than today²⁷. For our purposes, we expect the winter ice extent to have the main effect on deep-water ventilation, as stratification and biological productivity during summer independently limit the outgassing of CO_2 and uptake of O_2 . François *et al.*²⁸ used sediment $\delta^{15}\text{N}$ and opal data to estimate that increased stratification south of the APF during glacial times led to a 70% utilization of nutrients. This would be sufficient to remove $196 \mu\text{mol C kg}^{-1}$ ($0.7 \times 127 \times 2.2 \mu\text{mol}$ phosphate per kg) from the Antarctic surface waters, and thus prevent summertime CO_2 outgassing in our model. Furthermore, the extreme reduction in vertical mixing implied by the tenfold decrease in nutrient inputs estimated by François *et al.*²⁸ suggests that a large amount of ice remained in this region throughout the summer.

The results shown in Fig. 2 should be viewed as an indication of a potentially important mechanism in the ocean-atmosphere carbon system, rather than as an absolute prediction of the magnitude of the Antarctic sea-ice effect. However, we note that its magnitude is fairly robust with respect to variations in the assumed parameters. In addition to the sensitivities to ice-free area and F_a shown in Fig. 2a, the best-guess CO_2 difference changes by only -9 p.p.m. or $+10$ p.p.m. if the entrainment of low-latitude waters in AAIW is respectively doubled or reduced to zero. Although our model illustrates the behaviour of the ocean in the limit of no low-latitude deep-water upwelling, it is likely that there is some finite amount of diapycnal flow through the main thermocline in the real ocean. However, if we include a high-latitude sinking, low-latitude upwelling term of 10 Sv in our model, the total atmospheric CO_2 difference decreases by only 8 p.p.m. to a total of 59 p.p.m.

In addition, we have not tried to simulate increases in nutrient utilization in the subantarctic, yet in our model such changes could significantly affect the amount of CO_2 entering the deep ocean. If we reduce subantarctic surface nutrients after increasing Antarctic sea ice to generate an increase from 1.5 to $2.8 \text{ mol C m}^{-2} \text{ yr}^{-1}$ in subantarctic export production (slightly less than that proposed by François *et al.*²⁸), the CO_2 drawdown increases by 15 p.p.m. to a total of 82 p.p.m. A final perturbation to consider is that of temperature-driven solubility changes. After reducing surface temperatures as described above, our model predicts an additional CO_2 drawdown of 27 p.p.m., corresponding to a Harvardton Bear Index²³ of 0.2. The combined effects of ice cover, subantarctic productivity, and temperature simulated by our model are

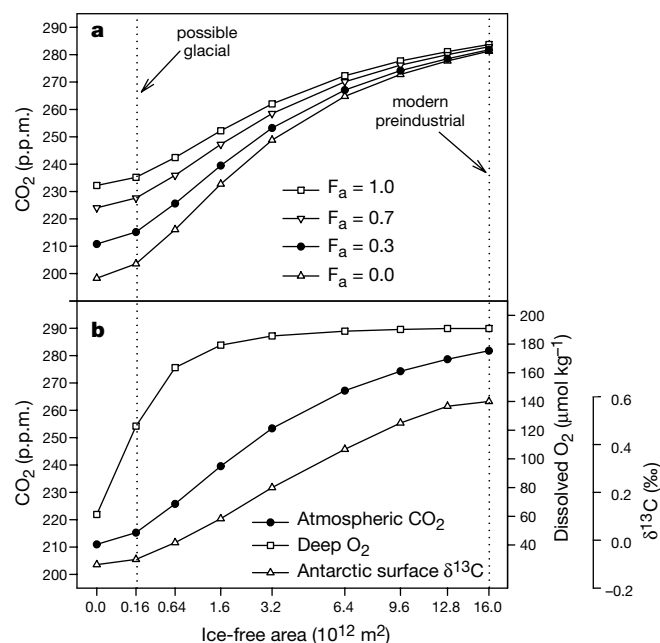


Figure 2 Steady-state model solutions for different ice coverages south of the Antarctic Polar Front (APF). We vary the exposed sea surface in the B and A boxes such that the same fraction of total surface is ice-free in each. The x-axis values represent the sum of exposed area in these two boxes, and are scaled by their square root to expand the left side of the plot. We use $1.6 \times 10^{13} \text{ m}^2$ as a modern estimate of the ice-free area south of the APF, and $1.6 \times 10^{11} \text{ m}^2$ to illustrate a possible glacial value. **a**, Atmospheric CO_2 using modern-preindustrial parameters and different fractions (F_a) of Antarctic surface waters entering the subantarctic box. Based on the evidence of Molinelli¹⁹ showing the importance of subsurface transport of Antarctic waters relative to vertical mixing as the subantarctic source of AAIW, we have chosen a relatively low F_a value of 0.3 to use as a best-guess case. **b**, Atmospheric CO_2 , deep O_2 , and Antarctic surface $\delta^{13}\text{C}$ using $F_a = 0.3$.

sufficiently large to ensure that an 80 p.p.m. decrease in glacial atmospheric CO₂ can still be attained even after including the counteracting effects of terrestrial biomass changes and salinity-driven solubility effects²³. Toggweiler²⁹ recently developed a model that similarly invokes a reduction in deep-water ventilation to explain the low glacial CO₂ levels. Whereas he generates reduced ventilation by decreasing the vertical exchange between deep and Antarctic surface waters, we suggest that reduced ventilation was driven by limitations to air–sea gas exchange imposed by increased sea ice in this region. □

Received 13 May 1999; accepted 18 January 2000.

- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Sarmiento, J. L. & Toggweiler, J. R. A new model for the role of the oceans in determining atmospheric pCO₂. *Nature* **308**, 621–624 (1984).
- Siegenthaler, U. & Wenk, T. Rapid atmospheric CO₂ variations and ocean circulation. *Nature* **308**, 624–626 (1984).
- Knox, F. & McElroy, M. B. Changes in atmospheric CO₂: Influence of the marine biota at high latitude. *J. Geophys. Res.* **89**, 4629–4637 (1984).
- Martin, J. H. Glacial–interglacial CO₂ change: The iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial–interglacial CO₂ changes. *Paleoceanography* **13**, 352–364 (1998).
- Opdyke, B. N. & Walker, J. C. G. Return of the coral reef hypothesis: Basin to shelf partitioning of CaCO₃ and its effect on atmospheric CO₂. *Geology* **20**, 733–736 (1982).
- Boyle, E. A. The role of vertical chemical fractionation in controlling Late Quaternary atmospheric carbon dioxide. *J. Geophys. Res.* **93**, 15701–15714 (1988).
- Archer, D. & Maier-Reimer, E. Effect of deep-sea sedimentary calcite preservation on atmospheric CO₂ concentration. *Nature* **367**, 260–263 (1994).
- Toggweiler, J. R. & Samuels, B. In *The Global Carbon Cycle* (ed. Heimann, M.) 333–366 (Springer, Berlin, 1993).
- Gnanadesikan, A. & Toggweiler, J. R. Constraints placed by silicon cycling on vertical exchange in general circulation models. *Geophys. Res. Lett.* **26**, 1865–1868 (1999).
- Charles, C. D. & Fairbanks, R. G. In *Geological History of the Polar Oceans: Arctic Versus Antarctic* (eds Bleil, U. & Thiede, J.) 519–538 (Kluwer Academic, Dordrecht, 1990).
- Boyle, E. A. Cadmium and δ¹³C paleochemical ocean distributions during the Stage 2 glacial maximum. *Annu. Rev. Earth Planet. Sci.* **20**, 245–287 (1992).
- Toggweiler, J. R. & Sarmiento, J. L. In *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 163–184 (American Geophysical Union, Washington DC, 1985).
- Marino, B. M., McElroy, M. B., Salawitch, R. J. & Spaulding, W. G. Glacial-to-interglacial variations in the carbon isotopic composition of atmospheric CO₂. *Nature* **357**, 461–466 (1992).
- Robbins, P. E. & Toole, J. M. The dissolved silica budget as a constraint on the meridional overturning circulation of the Indian Ocean. *Deep-Sea Res.* **44**, 879–906 (1997).
- Broecker, W. S. *et al.* How much deep water is formed in the Southern Ocean? *J. Geophys. Res.* **103**, 15833–15843 (1998).
- Tomczak, M. & Godfrey, J. S. *Regional Oceanography: An Introduction* 67–87 (Pergamon, Oxford, 1994).
- Molinelli, E. J. The Antarctic influence on Antarctic Intermediate Water. *J. Mar. Res.* **39**, 267–293 (1981).
- McCartney, M. S. in *A Voyage of Discovery* (ed. Angel, M.) 103–119 (Pergamon, Oxford, 1977).
- Berger, W. H. Deep-sea carbonate and the deglaciation preservation spike in pteropods and foraminifera. *Nature* **269**, 301–304 (1977).
- Broecker, W. S. An oceanographic explanation for the apparent carbon isotope–cadmium discordancy in the glacial Antarctic? *Paleoceanography* **8**, 137–139 (1993).
- Broecker, W. S. & Peng, T.-H. *Greenhouse Puzzles* 2nd edn (Eldigio, Palisades, New York, 1998).
- Guilderson, T. P., Fairbanks, R. G. & Rubinstone, J. L. Tropical temperature variations since 20,000 years ago: Modulating interhemispheric climate change. *Science* **263**, 663–665 (1994).
- Smith, H. J., Fischer, H., Wahlen, M., Mastropianni, D. & Deck, B. Dual modes of the carbon cycle since the Last Glacial Maximum. *Nature* **400**, 248–250 (1999).
- Cooke, D. W. & Hays, J. D. in *Antarctic Geoscience: Symposium on Antarctic Geology and Geophysics* (ed. Craddock, C.) 1017–1025 (University of Wisconsin Press, Madison, Wisconsin, 1982).
- Crosta, X., Pichon, J.-J. & Burckle, L. H. Reappraisal of Antarctic seasonal sea-ice at the Last Glacial Maximum. *Geophys. Res. Lett.* **26**, 1865–1868 (1999).
- François, R. *et al.* Contribution of Southern Ocean surface-water stratification to low atmospheric CO₂ concentrations during the last glacial period. *Nature* **389**, 929–935 (1997).
- Toggweiler, J. R. Variation of atmospheric CO₂ by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
- Broecker, W. S., Takahashi, T. & Takahashi, T. Sources and flow patterns of deep-ocean waters as deduced from potential temperature, salinity, and initial phosphate concentration. *J. Geophys. Res.* **90**, 6925–6939 (1985).

Acknowledgements

We thank R. Toggweiler, J. Sarmiento, C. Charles and J. Severinghaus for helpful discussions. This work was supported by the National Science Foundation and the Achievement Rewards for College Scientists Foundation.

Correspondence and requests for materials should be addressed to B.B.S. at the University of Colorado at Boulder (e-mail: britt@cmdl.noaa.gov).

Interferometric radar measurements of water level changes on the Amazon flood plain

Douglas E. Alsdorf*, John M. Melack*†, Thomas Dunne†, Leal A. K. Mertes*‡, Laura L. Hess* & Laurence C. Smith§

* Institute for Computational Earth System Science, University of California, Santa Barbara, California 93106, USA

† Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, USA

‡ Department of Geography, University of California, Santa Barbara, California 93106, USA

§ Department of Geography, University of California, Los Angeles, California 90095, USA

Measurements of water levels in the main channels of rivers, upland tributaries and floodplain lakes are necessary for understanding flooding hazards, methane production, sediment transport and nutrient exchange. But most remote river basins have only a few gauging stations and these tend to be restricted to large river channels. Although radar remote sensing techniques using interferometric phase measurements have the potential to greatly improve spatial sampling, the phase is temporally incoherent over open water and has therefore not been used to determine water levels. Here we use interferometric synthetic aperture radar (SAR) data^{1–3}, acquired over the central Amazon by the Space Shuttle imaging radar mission⁴, to measure subtle water level changes in an area of flooded vegetation on the Amazon flood plain. The technique makes use of the fact that flooded forests and floodplain lakes with emergent shrubs permit radar double-bounce returns from water and vegetation surfaces^{5,6}, thus allowing coherence to be maintained. Our interferometric phase observations show decreases in water levels of 7–11 cm per day for tributaries and lakes within ~20 km of a main channel and 2–5 cm per day at distances of ~80 km. Proximal floodplain observations are in close agreement with main-channel gauge records, indicating a rapid response of the flood plain to decreases in river stage. With additional data from future satellite missions, the technique described here should provide direct observations important for understanding flood dynamics and hydrologic exchange between rivers and flood plains.

Climatically driven, seasonal changes in river water levels (river stages) govern a wide range of hydrologic, geomorphological and ecological processes. Hydrologic modelling of the Amazon flood wave predicts discharge only along the main channel (the main stem), and suggests that up to 30% of mainstem flow exchanges with the flood plain⁷. On the basis of transport models, the annual sediment exchange between the main channel and flood plain is more than twice the flux through the most downstream river gauge at Obidos⁸. However, the models do not describe the sources or residence times of the floodplain water; these are key variables for measuring biological productivity and sedimentation. For example, observations of a small local catchment demonstrate that early in the water year, the lake contains nearly 70% river water whereas later in the water year, local runoff and other sources increase lake stage, preventing rising flood-stage river water from entering and exchanging with the lake⁹. Because very few of the ~8,000 Amazon floodplain lakes are gauged, the generality of these observations is unclear¹⁰.

Remotely sensed observations of the water surface provide an alternative to permanent gauging. Satellite radar altimetry promises a stage elevation accuracy of about 10 cm: but because altimetry is a profiling and not an imaging technique, it is applicable only to water

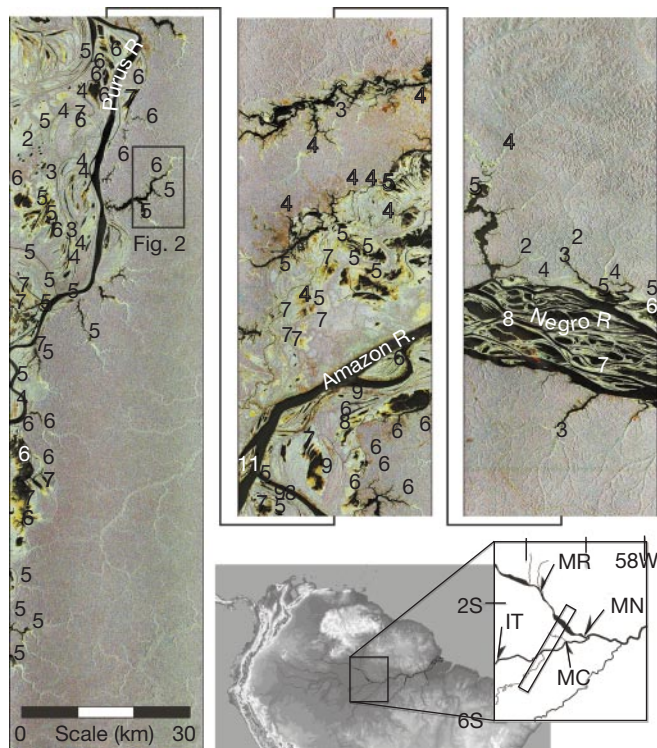


Figure 1 SIR-C swath and location map. Topographic map and close-up at lower right indicate the swath location (rectangle) relative to the surrounding gauging stations (MN, Manaus; MC, Manacapuru; IT, Itapeua; MR, Moura). The Solimões–Amazon River stretches from IT to MN and eastward, the Negro River extends from MN to MR and westward, and the Purus River joins the Solimões–Amazon in the southern half of the SIR-C swath. (2S and 6S indicate 2° and 6° S, 58W indicates 58° W.) The southern end of the 350-km-long SIR-C amplitude composite is the leftmost panel while the northern end is the rightmost panel. In the amplitude image, SIR-C C-band HH-polarization intensities are coloured red, L-HH are green, and L-HV are blue, thus yellows and greens within tributaries and floodplain lakes indicate strong, double-bounce L-HH returns marking inundated vegetation. Interferometrically determined local stage decreases in centimetres, from 9 to 10 October 1994, are noted by single-digit numbers, except the 11 cm drop near the confluence of the Purus with the Amazon river. For clarity, only half of the ~200 observed drops are noted. The amplitude composite is in SAR coordinates; thus the scale bar is approximate.

bodies greater than about two kilometres in width (its height accuracy is limited by orbital, range, and atmospheric errors^{11,12}). Imagery from space-borne platforms, such as passive microwave sensors, synthetic aperture radars and the Landsat thematic mapper, have been used to map the extent and timing of inundation^{5,13–16}, yet these methods do not directly measure stage elevation or change. Here we report that interferometric processing of SAR data from the SIR-C mission⁴ yields centimetre-scale measurements of stage change in floodplain lakes, floodplain channels (*paranas*), and upland tributaries (*igarapes*).

Radar backscatter intensity data from the second phase of the SIR-C mission were collected over the Amazon basin in October 1994, at X- (3.1 cm), C- (5.7 cm), and L-band (24.0 cm) wavelengths, and with horizontal and vertical polarizations. (By convention, for example, L-HV indicates a horizontally-sent vertically-received L-band radar pulse). Radar echoes at X- and C-band wavelengths backscatter from the tops of most forest canopies, while L-band echoes penetrate to the underlying ground or water surface: thus composite images from these simultaneously acquired radar echoes can be used to delineate vegetation types and floodplain inundation⁵ (Figs 1 and 2). Relative to nonflooded forests, increased backscatter intensities at L-HH are observed over flooded

forests, while compared to open water, C-HH backscattering increases over floating aquatic macrophytes and emergent shrubs in floodplain lakes. These flooded environments usually produce a two-bounce travel path of the radar pulse that includes interactions between the water surface and trunks of vegetation. Conversely, low backscattering is observed at both L and C bands over open water⁵. For example, the lake shown in Fig. 2 connects directly to the Purus River, a major tributary of the Solimões–Amazon River. Small tributaries drain the surrounding forested upland (for example, along profile B) and flood forests near their confluence with the lake (for example, along profile A).

Centimetre-scale topographic displacements resulting from earthquakes, inter-seismic strain accumulation, and flowing glaciers have been accurately measured using interferometric processing of SAR phase data^{1–3,17–21}. The processing method requires two SAR image acquisitions from identical (or nearly identical) viewing geometries before and after the displacement phenomenon; co-registration of the two images to a sub-pixel accuracy; and subtraction of the complex phase and amplitude values at each SAR image pixel. The value of the resulting interferometric phase at each pixel varies between $-\pi$ and $+\pi$, and is primarily a function of the distance between the radar antenna positions during acquisition (that is, the Shuttle baseline), topographic relief, topographic displacement, and the degree of correlation between the individual scattering elements that comprise each pixel location (that is, coherence^{22,23}). As expected for C-band wavelengths^{24–27}, the scattering elements of the tree canopy (that is, leaves and branches easily moved by gentle breezes) completely decorrelated in the 24 hours between SIR-C acquisitions, whereas at L-band, the elements were typically comprised of earth and immobile trunks of vegetation which enabled strong phase-correlation (see profiles in Fig. 2).

The most unexpected result of the interferometric processing is the strong, L-band coherence (>0.5) maintained over regions marked by flooded forest and floodplain lakes with tree canopies and emergent shrubs (profiles shown in Fig. 2). Radar pulse interactions with open water are specular (that is, reflect away from the antenna); thus L-band coherence over open rivers and lakes was low, averaging 0.26. However, the strong coherence of tributaries within flooded forests and of floodplain lakes with vegetation enabled interferometric phase measurements of the earth, vegetation trunks, and water surfaces that returned the radar echoes. In these environments, the higher-amplitude L-HH returns compared to L-HV⁵ produced stronger correlation, thus leading to more reliable interferometric phase measurements.

The observed interferometric phase measured both the local topography and also any surface displacements that occurred during the 24 hours separating the two SIR-C acquisitions. To separate these components, either a synthetic interferogram based on a digital elevation model (DEM) or additional SAR interferograms free of displacement phenomena can be subtracted from the observed phase³. Unfortunately, neither a high-resolution DEM nor a third SAR acquisition were available. Instead, the short, 9–10 October Shuttle baseline (perpendicular component of only 63 m) enables discrimination of the topographic and displacement components of the total measured phase. Using this baseline, the amount of topographic relief measured from $-\pi$ to $+\pi$ at the L-band wavelength is 275 m (refs 24, 25). Since the relief throughout the study area is known to be less than ~65 m, only ~1/4 of a cycle ($\sim 1/2\pi$ or less) of the observed interferometric phase can be attributed to topography. Indeed, for the imaged floodplain and upland areas, phase values of only 0.0 to -0.5 radians (blue pixels in Fig. 2) are found in the interferogram, indicating low regional relief.

Although the short baseline prohibits accurate measurements of Amazon basin topography²⁷, its orientation enables simple physiographic comparisons. For example, the phase values in profile B of Fig. 2 reach peak values where crossing an upstream valley (given the SIR-C baseline orientation, phase peaks are expected over topographic

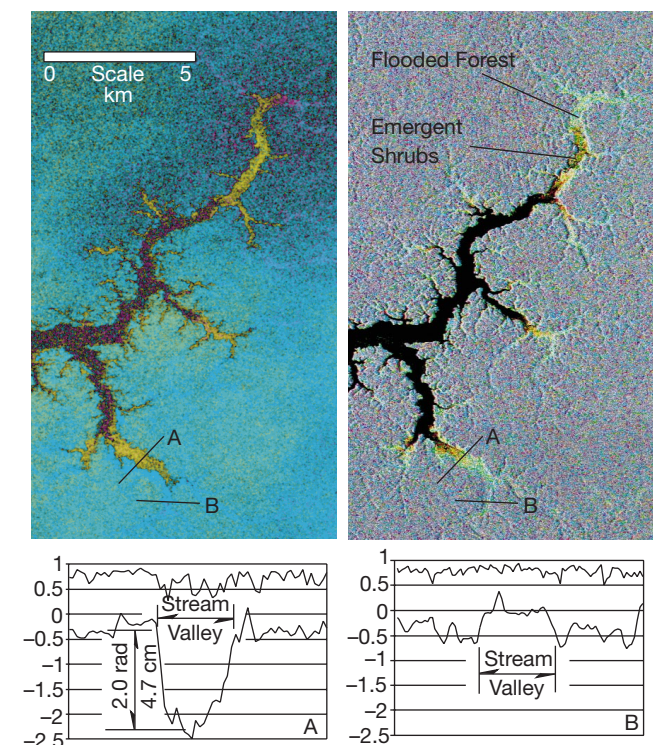


Figure 2 SIR-C L-HH flattened interferogram (left) and amplitude composite (right). See Fig. 1 for location of this area along the Purus River, a tributary of the Amazon–Solimões River. A and B mark profiles of interferometric phase (bottom line in both profiles; in this case, units on vertical axes are radians) and corresponding interferometric coherence (top line in both profiles; in this case, units are interferometric correlation). Using the red–green–blue colour scheme of Fig. 1 on the amplitude composite, flooded forests in the upper tributary reaches are bright yellow and green while open water is black. Tributaries marked by mottled-greenish pixels indicate emergent shrubs. Flooded forests and emergent shrubs are marked by coherent interferometric phase values of about -2.5 radians (yellow pixels in the interferogram). Phase image brightness is a function of the local L-HH backscatter intensity.

valleys). Yet, unexpectedly, the phase values along profile A reach their lowest values where crossing the same valley further downstream. This difference cannot be ascribed to poor correlation (correlation values for both plots are almost all greater than 0.5) nor to a topographic ridge (the phase trough is ~ 4 times greater in A than the phase peak in B, thus any possible topographic ridge would encompass ~ 4 times the relief of the valley—the amplitude image does not show such a ridge). Phase differences related to atmospheric water vapour can also be discounted because they are usually longer in wavelength and spatially distinct from phase signatures related to tributaries and floodplain lakes. Therefore, the phase trough must indicate a surface displacement away from the Shuttle. Given the spatial distribution of the phase trough, a decrease in the elevation of the water surface that occurred during the 24 hours separating the two SIR-C acquisitions is the most likely displacement phenomenon. As the water surface elevation change is purely vertical, the 2.0-radian trough can be projected from the SAR line-of-sight direction to yield a stage decrease of nearly 5 cm ($\Delta z = \lambda \varphi / [4\pi \cos(\text{look angle})]$, where $\lambda = 24.0$ cm, φ is the displacement phase, the look angle is 35.2° , and Δz is the vertical displacement).

Using the method described above, the change in stage for vegetation-covered floodplain lakes and tributaries in flooded forests was measured throughout the SIR-C interferometric swath (Fig. 1). Drainage path distances were measured from each water body to the closest mainstem river, that is, the Solimões–Amazon, Negro or Purus rivers. The largest drops of 7–11 cm were measured

in floodplain water bodies within ~ 20 km of either the Solimões–Amazon or Purus Rivers or within the Negro archipelago. For comparison, the Manaus gauge (Fig. 1) recorded a drop of 12 cm in water level from 9 to 10 October 1994, and an average daily decrease of 12 cm in the previous 7 days. The Manacapuru, Itapeua and Moura gauges (Fig. 1) recorded decreases of 7, 5 and 6 cm, respectively, while averages over the previous 7 days were 11, 7 and 11 cm, respectively.

Interferometric measurements of decreasing floodplain water levels indicate that the proximal flood plain was rapidly evacuating water to mainstem channels during the SIR-C acquisitions, in response to the 11 cm drop in channel stage. At distances greater than ~ 30 km, where the flood plain includes input from surrounding upland tributaries and lakes, drops were 2–6 cm. Although these distal water bodies ultimately discharge to a mainstem river, it is likely that many days are necessary for exchange. Consecutive daily decreases of mainstem water levels since early July 1994 may have worked to decrease distal water levels while local catchment inputs (precipitation, groundwater flow and runoff) worked to increase water levels. For comparison, tributaries and lakes in upland areas that drain directly to the Solimões–Amazon or Purus Rivers also had interferometrically measured drops of 2–6 cm, with the largest drops occurring within 20 km of either mainstem river. Viewed collectively, the dominant trend in the SIR-C observations is that drops in water level diminish with increasing distance from a mainstem river. Proximal floodplain waters are quickly evacuated to mainstem rivers in response to decreased river discharge, even in low-lying areas of the Amazon. This indicates the importance and rapidity of hydrologic transfer from flood plain to river channel during recession flow.

These observations from the SIR-C mission provide, we believe for the first time, a spatial image of centimetre-scale variations in floodplain water level response to changing river discharge; they imply that future satellite L-HH band SAR missions are capable of yielding interferometric measurements of stage change over the entire Amazon flood plain. Furthermore, our observations show that planning and acquiring L-HH interferometric radar data over other large river systems—such as the Congo of central Africa and smaller systems such as the Alligator rivers near Darwin, Australia—should yield spatial and temporal measurements of stage changes where vegetation characteristics permit interferometric correlation to be maintained. Thus, with a broader application to global wetlands, our work suggests that the complex dynamics of floods, including the timing and rates of hydrologic exchange between channels and flood plains, can be remotely measured with greatly improved resolution. □

Received 16 July 1999; accepted 17 January 2000.

- Goldstein, R. M., Engelhardt, H., Kamb, B. & Frolich, R. M. Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* **262**, 1525–1530 (1993).
- Massonnet, D. *et al.* The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* **364**, 138–142 (1993).
- Zebker, H. A., Rosen, P. A., Goldstein, R. M., Gabriel, A. & Werner, C. L. On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. *J. Geophys. Res.* **99**, 19617–19634 (1994).
- Stofan, E. R., *et al.* Overview of results of Spaceborne Imaging Radar-C, X-band synthetic aperture radar (SIR-C/X-SAR). *IEEE Trans. Geosci. Remote Sensing* **33**, 817–828 (1995).
- Hess, L. L., Melack, J. M., Filoso, S. & Wang, Y. Delineation of inundated area and vegetation along the Amazon floodplain with SIR-C synthetic aperture radar. *IEEE Trans. Geosci. Remote Sensing* **33**, 896–904 (1995).
- Wang, Y., Hess, L. L., Filoso, S. & Melack, J. M. Understanding the radar backscattering from flooded and nonflooded Amazonian forests: Results from canopy backscatter modeling. *Remote Sensing Environ.* **54**, 324–332 (1995).
- Richey, J. E. *et al.* Sources and routing of the Amazon river flood wave. *Glob. Biogeochem. Cycles* **3**, 191–204 (1989).
- Dunne, T., Mertes, L. A. K., Meade, R. H., Richey, J. E. & Forsberg, B. R. Exchanges of sediment between the flood plain and channel of the Amazon River in Brazil. *GSA Bull.* **110**, 450–467 (1998).
- Lesack, L. F. W. & Melack, J. M. Flooding hydrology and mixture dynamics of lake water derived from multiple sources in an Amazon floodplain lake. *Wat. Resour. Res.* **31**, 329–345 (1995).
- Melack, J. M. & Forsberg, B. R. in *The Biogeochemistry of the Amazon Basin and its Role in a Changing World* (eds McClain, M. E., Victoria, R. L., Richey, J. E.) (Oxford Univ. Press, New York, in the press).
- Birkett, C. M. The contribution of TOPEX/POSEIDON to the global monitoring of climatically

- sensitive lakes. *J. Geophys. Res.* **100**, 25179–25204 (1995).
12. Birkett, C. M. Contribution of the TOPEX/Poseidon altimeter to the global monitoring of large rivers and wetlands. *Wat. Resour. Res.* **34**, 1223–1239 (1998).
 13. Mertes, L. A. K. Documentation and significance of the perirheic zone on inundated floodplains. *Wat. Resour. Res.* **33**, 1749–1762 (1997).
 14. Smith, L. C. Satellite remote sensing of river inundation area, stage, and discharge: A review. *Hydrol. Processes* **11**, 1427–1439 (1997).
 15. Mertes, L. A. K. et al. Spatial patterns of hydrology, geomorphology, and vegetation on the floodplain of the Amazon River in Brazil from a remote sensing perspective. *Geomorphology* **13**, 215–232 (1995).
 16. Sippel, S. J., Hamilton, S. K., Melack, J. M. & Novo, E. M. M. Passive microwave observations of inundation area and the area/stage relation in the Amazon River floodplain. *Int. J. Remote Sensing* **19**, 3055–3074 (1998).
 17. Peltzer, G. & Rosen, P. Surface displacement of the 17 May 1993 Eureka Valley, California, earthquake observed by SAR interferometry. *Science* **268**, 1333–1336 (1995).
 18. Wicks, C., Thatcher, W. & Dzurisin, D. Migration of fluids beneath Yellowstone caldera inferred from satellite radar interferometry. *Science* **282**, 458–462 (1998).
 19. Burgmann, R., Fielding, E. & Sukhatme, J. Slip along the Hayward fault, California, estimated from space-based synthetic aperture radar interferometry. *Geology* **26**, 559–562 (1998).
 20. Rignot, E. J., Gogineni, S. P., Krabill, W. B. & Ekholm, S. North and northeast Greenland ice discharge from satellite radar interferometry. *Science* **272**, 934–937 (1997).
 21. Joughin, I., Kwok, R. & Fahnestock, M. Estimation of ice-sheet motion using satellite radar interferometry: Method and error analysis with application to Humboldt Glacier, Greenland. *J. Glaciol.* **42**, 564–575 (1996).
 22. Li, F. K. & Goldstein, R. M. Studies of multibaseline spaceborne interferometric synthetic aperture radars. *IEEE Trans. Geosci. Remote Sensing* **28**, 88–97 (1990).
 23. Zebker, H. A. & Villaseñor, J. Decorrelation in interferometric radar echoes. *IEEE Trans. Geosci. Remote Sensing* **30**, 950–959 (1992).
 24. Rosen, P. A., Hensley, S., Zebker, H. A., Webb, F. H. & Fielding, E. J. Surface deformation and coherence measurements of Kilauea Volcano, Hawaii, from SIR-C radar interferometry. *J. Geophys. Res.* **101**, 23109–23125 (1996).
 25. Massonnet, D. & Rabaute, T. Radar interferometry: Limits and potential. *IEEE Trans. Geosci. Remote Sensing* **31**, 455–464 (1993).
 26. Zebker, H. A., Rosen, P. A., Hensley, S. & Mouginiis-Mark, P. J. Analysis of active lava flows on Kilauea volcano, Hawaii, using SIR-C radar correlation measurements. *Geology* **24**, 495–498 (1996).
 27. Rignot, E. Dual-frequency interferometric SAR observations of a tropical rain-forest. *Geophys. Res. Lett.* **23**, 993–996 (1996).

Acknowledgements

We thank H. Zebker, C. Birkett and J. Ridley for comments on the manuscript. The Cornell Theory Center provided the computer resources necessary for SAR processing. Gamma Remote Sensing Research and Consulting provided SAR technical advice. This work was supported by NASA.

Correspondence and requests for materials should be addressed to D.E.A. (e-mail: alsdorf@icess.ucsb.edu).

Delayed biological recovery from extinctions throughout the fossil record

James W. Kirchner* & Anne Weil†

* Department of Geology and Geophysics, University of California, Berkeley, California, 94720-4767, USA

† Department of Biological Anthropology and Anatomy, Duke University, Durham, North Carolina 27708-0383, USA

How quickly does biodiversity rebound after extinctions? Palaeobiologists have examined the temporal, taxonomic and geographic patterns of recovery following individual mass extinctions in detail^{1–5}, but have not analysed recoveries from extinctions throughout the fossil record as a whole. Here, we measure how fast biodiversity rebounds after extinctions in general, rather than after individual mass extinctions, by calculating the cross-correlation between extinction and origination rates across the entire Phanerozoic marine fossil record. Our results show that extinction rates are not significantly correlated with contemporaneous origination rates, but instead are correlated with origination rates roughly 10 million years later. This lagged correlation persists when we remove the ‘Big Five’ major mass extinctions, indicating that recovery times following mass

extinctions and background extinctions are similar. Our results suggest that there are intrinsic limits to how quickly global biodiversity can recover after extinction events, regardless of their magnitude. They also imply that today’s anthropogenic extinctions will diminish biodiversity for millions of years to come.

A key component of biotic recovery is the time lag between episodes of rapid extinction and subsequent periods of rapid origination. Originations rebuild biodiversity, and origination rates are commonly assumed to peak when ecosystems have recovered sufficient diversity to inhibit further diversification⁶. Thus, the elapsed time between extinction rate peaks and origination rate peaks is one measure of the recovery time (Fig. 1). We can estimate the average time lag for the whole fossil record using the cross-correlation between extinctions and originations, which measures how closely the two time series resemble each other, when one is shifted forwards or backwards by a specified interval. For regularly spaced time series, the cross-correlation function could be calculated as

$$r_{EO}(k) = \frac{\sum (E_i - \bar{E})(O_j - \bar{O})}{\sqrt{\sum (E_i - \bar{E})^2} \sqrt{\sum (O_j - \bar{O})^2}}, \quad j = i + k \quad (1)$$

where E and O are the extinction and origination time series, and $r_{EO}(k)$ is their cross-correlation when originations lag extinctions by k steps. This direct approach cannot be applied to the fossil record, because its stratigraphic boundaries are unevenly spaced in time. Nor should one simply even out the spacing by interpolating within each stratigraphic interval, because this introduces artefactual correlation among the interpolated points^{7,8}. Instead, we use equation (1) to calculate the correlation between all pairs of points E_i and O_j whose separations in time, $t_j - t_i$, fall within 5 Myr bins of lag time (rather than points separated by fixed numbers of steps); this yields a geostatistical approximation to the cross-correlation function⁹ (Fig. 2a, b). A second approach to calculating the cross-correlation between two unevenly spaced time series is through their Fourier transforms^{10,11}:

$$r_{EO}(\tau) = FT^{-1}(FT\{E\}FT^*\{O\}) \quad (2)$$

where FT , FT^* and FT^{-1} denote the Fourier transform, its complex conjugate and its inverse, respectively, and $r_{EO}(\tau)$ is the cross-correlation between E and O at a lag of τ Myr (Fig. 2c, d). We use the Lomb–Scargle Fourier transform^{11–14} to calculate $FT\{E\}$ and $FT^*\{O\}$ directly from the unevenly spaced fossil data, without interpolation. The Lomb–Scargle algorithm has similar statistical properties, when applied to unevenly spaced data, to those of conventional Fourier transform algorithms when applied to evenly spaced data¹⁴.

Our source data are Sepkoski’s compilations of fossil marine animal genera¹⁵ and families¹⁶, with revisions through 1997. Because long-term drift could obscure the cross-correlations that we seek to analyse, we subtracted the long-term trends (shown as dotted lines in Fig. 1b, c) from the extinction and origination time series before analysis.

The resulting cross-correlation functions (Fig. 2) show that extinctions and originations are not significantly correlated over short lag periods, indicating that, on average, extinctions do not trigger immediate evolutionary rebounds. Instead, the cross-correlation is strongest when originations lag extinctions by roughly 10 Myr. This indicates that the average interval between extinction peaks and origination peaks, and thus the average recovery time from extinctions, is about 10 Myr across the fossil record. The statistical significance of these cross-correlations—that is, the chance of correlations this strong arising by chance at any lag, from –15 to 35 Myr—is $P < 0.05$ for all but one of the 16 cases shown in Fig. 2 (see Supplementary Information). The peak cross-correlation occurs at similar lags with either calculation method, and in both the genus and family data sets, indicating that this result

is robust.

To test whether these results are driven by the mass extinctions, we repeated our analysis after excluding the 'Big Five' extinction peaks (Fig. 1) from the data set. The resulting cross-correlations (heavy dotted lines, Fig. 2) are weaker, but still have their strongest

positive values when originations lag extinctions by roughly 10 Myr. We also rank-transformed the extinction and origination rates to obtain a cross-correlation function, analogous to the Spearman rank correlation, that minimizes the influence of extreme values associated with mass extinctions. The results of this analysis (dashed

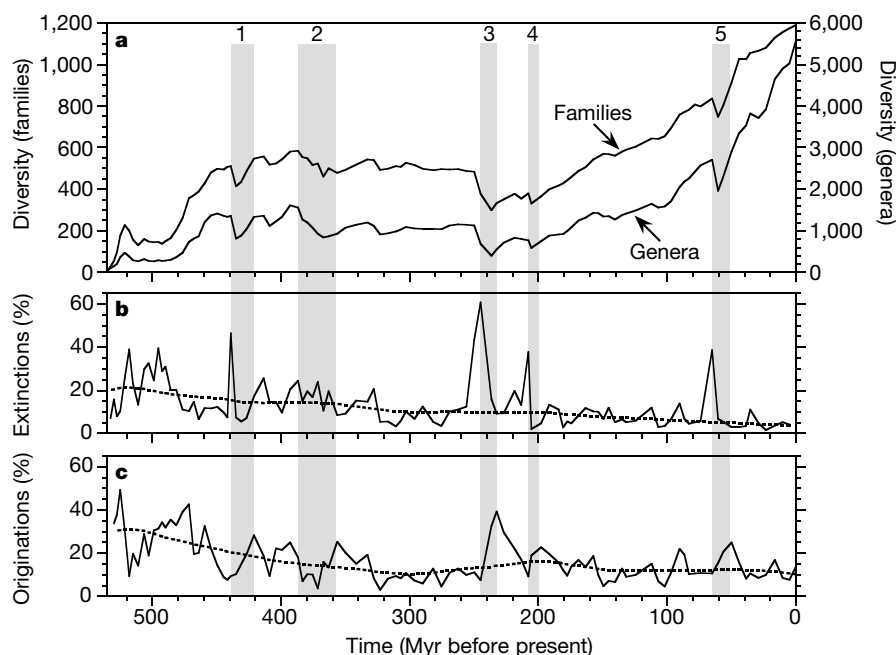


Figure 1 The fossil record of marine animal biodiversity. Standing diversity of genera and families through the Phanerozoic (a), and corresponding percentages of extinction (b) and origination (c) of genera in each stratigraphic interval. Shaded bands highlight recovery intervals (between extinction rate peaks and subsequent origination rate peaks) for the

'Big Five' mass extinctions: end-Ordovician (1), late Devonian (2), end-Permian (3), end-Triassic (4) and Cretaceous-Tertiary (5). Dotted lines in b and c show the long-term trends (estimated using LOWESS, a robust curve-fitting technique¹⁹) that are subtracted from extinction and origination time series before calculating cross-correlations.

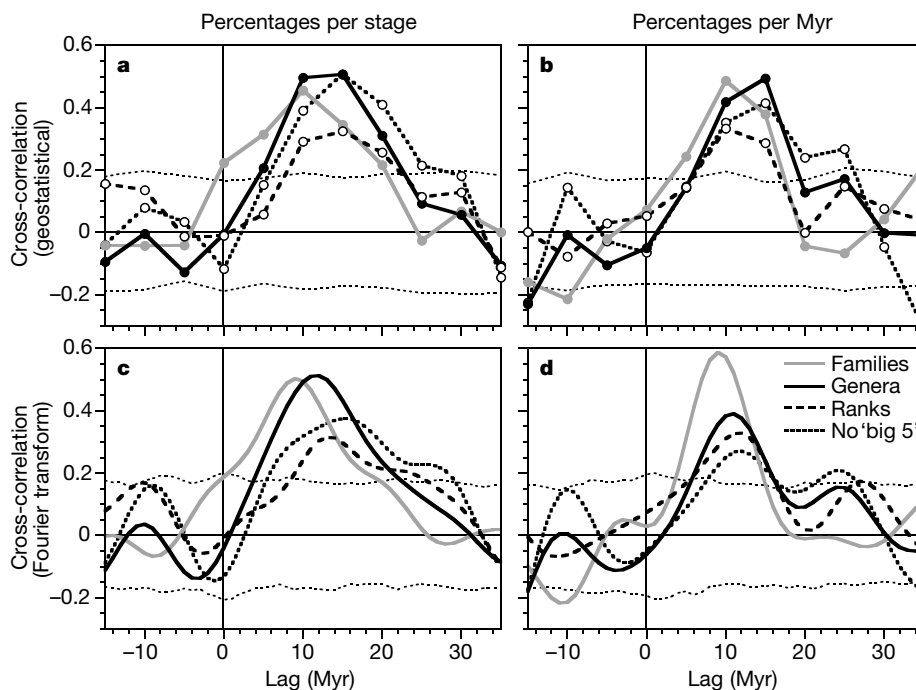


Figure 2 Cross-correlation between extinctions and originations. Cross-correlation functions calculated by geostatistical (a, b) and Fourier transform (c, d) methods, from percentage rates of extinction and origination per stratigraphic interval (a, c) and per million years (b, d). Cross-correlations are shown for genera (black lines), families (grey lines), rank-transformed genera (dashed lines) and genera with the 'Big Five' major mass

extinctions removed (heavy dotted lines). Fine dotted lines indicate upper and lower 5% confidence bounds for uncorrelated time series, calculated from cross-correlations of 1,000 random shuffles of the original data. Positive lags indicate originations lagging extinctions.

lines, Fig. 2) are similar to those obtained by deleting the 'Big Five' events. Considered together, these results indicate that recovery times on the order of 10 Myr, previously noted for mass extinctions^{3–5}, are also characteristic of background extinctions.

The incompleteness of the fossil record implies that last occurrences of fossils, and thus apparent extinctions, are earlier than actual dates of extinction¹⁷ (the Signor–Lipps effect); it also implies that first occurrences of fossils, and thus apparent originations, are later than actual dates of origination⁵ (the 'Sppil–Rongis' effect). Could these two artefacts create lagged correlations similar to those that we observe? We tested this possibility by assuming that originations at each stage boundary were exactly equal to extinctions—thus creating perfectly correlated records, with no lag—and then spreading apparent extinctions backwards in time and apparent originations forwards in time, as one would expect from incomplete sampling (see Methods). Figure 3 shows that even a very incomplete fossil record, with average intervals of 5 or 10 Myr between fossil discoveries, would not produce lagged correlations similar to those we observed: smearing the fossil record stretches the cross-correlation function toward longer lags, but does not shift its peak. Therefore, the lagged cross-correlation in the real fossil data cannot be explained by incomplete sampling of the fossil record.

As similar recovery intervals appear to follow both mass extinctions and background extinctions, the timescale of recovery may be determined not by the magnitude of individual extinctions, but instead by the internal dynamics of the diversification process. Species do not merely occupy ecological niches; they also serve as evolutionary starting points for radiation into additional niches, and they themselves constitute niches for their predators, parasites and symbionts. Consequently, as extinctions eliminate species they also destroy niches, and may thus reduce biodiversity without creating many evolutionary opportunities. Instead, new opportunities for radiation may arise mostly from diversification itself, which, in creating new taxa, creates new niches and new evolutionary pathways into existing niches. This implies that, after extinctions, origination rates should initially be low, accelerating only as diversification creates niche space, and finally peaking when ecosystem structure is sufficiently developed to slow further diversification. The duration of this process, and thus the recovery timescale, will depend on the structure of the post-extinction ecosystem, which will be configured differently than the pre-extinction ecosystem. If the structure of the new ecosystem depends on biogeographic and taxonomic factors², the duration of the recovery process may be largely independent of extinction magnitude. Thus, although mass extinctions are larger than background

extinctions, we should not expect their recoveries to be proportionally or consistently longer.

Multimillion-year recovery intervals are a previously unrecognized general property of the fossil record, not a phenomenon associated only with mass extinctions. Our results suggest that there are intrinsic 'speed limits' that regulate recovery from small extinctions as well as large ones. Thus, today's anthropogenic extinctions are likely to have long-lasting effects, whether or not they are comparable in scope to the major mass extinctions. Even if *Homo sapiens* survives several million more years, it is unlikely that any of our species will see biodiversity recover from today's extinctions. □

Methods

Source data

Sepkoski's timescale aggregates some stratigraphic stages and subdivides others, yielding 106 intervals, 2.5 to 12.5 Myr in length, from the Cambrian through the Pleistocene. To minimize Lagerstätten and monographic effects¹⁵, we excluded all taxa that occur in only one stratigraphic interval. We analysed originations and extinctions as percentages per stratigraphic interval (originations or extinctions divided by total diversity in each interval) and percentages per unit time (originations or extinctions divided by total diversity and interval length)¹⁸. Our family cross-correlations (grey lines, Fig. 2) exclude the Cambrian, because its low initial familial diversity yields anomalously high percentages of extinctions and originations. We assigned originations in each interval to the stratigraphic boundary that begins it, and extinctions to the stratigraphic boundary that ends it; this approach is conservative because it minimizes the time by which originations lag extinctions. To the extent that faunal turnover occurred within (rather than between) stratigraphic intervals, and thus extinctions and originations within each interval were highly correlated, our procedure would yield an artefactual negative lag (that is, extinctions lagging originations), the opposite of what we observe. Assigning originations and extinctions to the midpoint of each stratigraphic interval would lengthen the lag between extinctions and subsequent originations by roughly the average interval length (5 Myr), but would not significantly change the shape of the cross-correlation functions (see Supplementary Information).

Simulating incomplete sampling

If fossil discoveries from a given genus or family are randomly distributed through time (a Poisson process), the interval Δt between the actual time of origination and the first fossil found (or between the actual time of extinction and the last fossil found) will follow an exponential probability distribution:

$$p(\Delta t) = \frac{1}{T} e^{-\Delta t/T} \quad (3)$$

where T is the average interval between individual fossil discoveries from each genus or family. The apparent extinction or origination time series can be estimated by convolving equation (3) with the actual time series of extinctions or originations.

Received 30 September 1999; accepted 5 January 2000.

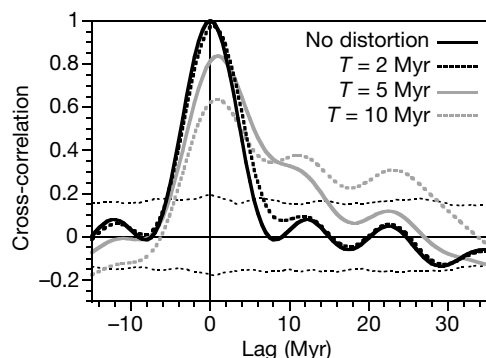


Figure 3 Effect of incomplete sampling. Cross-correlation functions (calculated by Fourier transform method) are shown for synthetic, perfectly correlated extinction and origination time series, with and without distortion by simulated Signor–Lipps and 'Sppil–Rongis' effects (see Methods). Curves are shown for different degrees of distortion, corresponding to different average intervals T between fossil discoveries in each taxon. Backwards-smearing of extinctions and forwards-smearing of originations do not shift the peak cross-correlation, but instead stretch the cross-correlation function toward longer lags.

- Hallam, A. Why was there a delayed radiation after the end-Palaeozoic extinctions? *Hist. Biol.* **5**, 257–262 (1991).
- Jablonski, D. Geographic variation in the molluscan recovery from the end-Cretaceous extinction. *Science* **279**, 1327–1330 (1998).
- Jablonski, D. in *Extinction Rates* (eds Lawton, J. H. & May, R. M.) 25–44 (Oxford Univ. Press, New York, 1995).
- Erwin, D. H. The end and the beginning: recoveries from mass extinctions. *Trends Ecol. Evol.* **13**, 344–349 (1998).
- Sepkoski, J. J. Jr Rates of speciation in the fossil record. *Phil. Trans. R. Soc. Lond. B* **353**, 315–326 (1998).
- Walliser, O. H. in *Global Events and Event Stratigraphy in the Phanerozoic* (ed. Walliser, O. H.) 7–19 (Springer, Berlin, 1996).
- Kirchner, J. W. & Weil, A. No fractals in fossil extinction statistics. *Nature* **395**, 337–338 (1998).
- Schulz, M. & Stettin, K. SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series. *Comput. Geosci.* **23**, 929–945 (1997).
- Agterberg, F. P. in *Geostatistics* (ed. Merriam, D. E.) 113–141 (Plenum, New York, 1970).
- Bracewell, R. N. *The Fourier Transform and its Applications* 3rd edn (McGraw Hill, Boston, 2000).
- Scargle, J. D. Studies in astronomical time series analysis. III. Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *Astrophys. J.* **343**, 874–887 (1989).
- Lomb, N. R. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **29**, 447–462 (1976).
- Press, W. H. & Teukolsky, S. A. Search algorithm for weak periodic signals in unevenly spaced data. *Comput. Phys.* **2**, 77–82 (1988).
- Scargle, J. D. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).
- Sepkoski, J. J. Jr in *Global Events and Event Stratigraphy in the Phanerozoic* (ed. Walliser, O. H.) 35–51 (Springer, Berlin, 1996).
- Sepkoski, J. J. Jr Ten years in the library: new data confirm paleontological patterns. *Paleobiology* **19**, 43–51 (1993).
- Signor, P. W. & Lipps, J. H. Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. *Geol. Soc. Am. Special Paper* **190**, 291–296 (1982).

18. Gilinsky, N. L. in *Analytical Paleobiology* (eds Gilinsky, N. L. & Signor, P. W.) 157–174 (The Paleontological Society, Knoxville, Tennessee, 1991).
 19. Cleveland, W. S. & McGill, R. The many faces of a scatterplot. *J. Am. Stat. Assoc.* **79**, 807–822 (1984).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We are indebted to the late J. Sepkoski for his fossil databases and his encouragement, and we thank M. Foote and D. Erwin for comments on the manuscript. Our work was supported by grants from the University of California and the NSF to J.W.K.

Correspondence and requests for materials should be addressed to J.W.K. (e-mail: kirchner@seismo.berkeley.edu)

Simple rules yield complex food webs

Richard J. Williams & Neo D. Martinez

Romberg Tiburon Center, Department of Biology, San Francisco State University, PO Box 855, Tiburon, California 94920, USA

Several of the most ambitious theories in ecology^{1–14} describe food webs that document the structure of strong and weak trophic links⁹ that is responsible for ecological dynamics among diverse assemblages of species^{4,11–13}. Early mechanism-based theory asserted that food webs have little omnivory and several properties that are independent of species richness^{1–4,6}. This theory was overturned by empirical studies that found food webs to be much more complex^{5,7–9,14–18}, but these studies did not provide mechanistic explanations for the complexity⁹. Here we show that a remarkably simple model fills this scientific void by successfully predicting key structural properties of the most complex and comprehensive food webs in the primary literature. These properties include the fractions of species at top, intermediate and basal trophic levels, the means and variabilities of generality, vulnerability and food-chain length, and the degrees of cannibalism, omnivory, looping and trophic similarity. Using only two empirical parameters, species number and connectance, our 'niche model' extends the existing 'cascade model'^{3,19} and improves its fit ten-fold by constraining species to consume a contiguous sequence of prey in a one-dimensional trophic niche²⁰.

We compare the abilities of two earlier models, the random and cascade models^{3,19}, and our new niche model to predict a dozen properties for each of seven food webs. The parameters of all models are set to synthesize webs with the empirically observed species number and connectance level. We compare model predictions with the largest and highest-quality empirical food webs that include autotrophs and were originally documented to study food web structure comprehensively (Table 1). Three are from freshwater habitats: Skipwith Pond, Little Rock Lake and Bridge Brook Lake;

two are from habitats at freshwater-marine interfaces: Chesapeake Bay and Ythan Estuary; and two are from terrestrial habitats: Coachella Valley and the island of St Martin.

Throughout this work, 'species' refers to trophic species, which are functional groups of taxa that share the same predators and prey in a food web³. 'Trophic species' is a widely accepted^{3,4,8,14,17,18} and sometimes criticized convention^{5,14} within structural food-web studies that reduces methodological biases in the data^{3,4,8}. A matrix with S rows and columns represents a food web with S species. Element a_{ij} is 1 if species j consumes species i and 0 if not. There are S^2 possible and L actual links. Directed connectance¹⁷ (C) equals L/S^2 .

In the random model^{3,19}, any link among S species occurs with the same probability (P) equal to C of the empirical web. This creates webs as free as possible from biological structuring while maintaining the observed S and C . The cascade model^{3,19} assigns each species a random value drawn uniformly from the interval $[0,1]$ and each species has probability $P = 2CS/(S-1)$ of consuming only species with values less than its own. This pecking order helps to explain species richness among trophic levels³ but underestimates interspecific trophic similarity¹⁹ and overestimates food-chain length and number in larger webs^{3,18}. The niche model (Fig. 1) similarly assigns each species a randomly drawn 'niche value'. The species are then constrained to consume all prey species within one range of values whose randomly chosen centre is less than the consumer's niche value. The single range adds a previously discussed²⁰ community-level contiguity of niche space to the cascade model by causing species with similar niche values to share consumers frequently within the community. The placement of the niche partially relaxes the cascade hierarchy by allowing up to half a consumer's range to include species with niche values higher than the consumer's value. All three models incorporate substantial stochastic variability along with dependence on S and C .

Twelve properties of each empirical and model web are measured (see Methods):

(i–iii) Species types^{1–8,14–18,21}: the fractions of top (T , species with no predators), intermediate (I , species with both predators and prey) and basal (B , species with no prey) species.

(iv, v) The standard deviations (s.d.) of generality¹⁴ ($GenSD$) and vulnerability¹⁴ ($VulSD$) quantify the respective variabilities of species' normalized prey (G_i) and predator (V_i) counts:

$$G_i = \frac{1}{L/S} \sum_{j=1}^S a_{ji} \quad V_i = \frac{1}{L/S} \sum_{j=1}^S a_{ij}$$

Normalizing with L/S makes s.d. comparable across different webs by forcing mean G_i and V_i to equal 1.

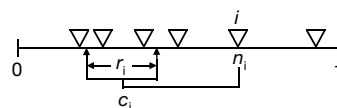


Figure 1 Diagram of the niche model. Each of S species (for example, $S = 6$, each shown as an inverted triangle) is assigned a 'niche value' parameter (n_i) drawn uniformly from the interval $[0,1]$. Species i consumes all species falling in a range (r_i) that is placed by uniformly drawing the centre of the range (c_i) from $[r_i/2, n_i]$. This permits looping and cannibalism by allowing up to half of r_i to include values $\geq n_i$. The size of r_i is assigned by using a beta function to randomly draw values from $[0,1]$ whose expected value is $2C$ and then multiplying that value by n_i [expected $E(n_i) = 0.5$] to obtain the desired C . A beta distribution with $\alpha = 1$ has the form $f(x) = \beta(1-x)^{\beta-1}$, $0 < x < 1$, 0 otherwise, and $E(x) = 1/(1+\beta)$. In this case, $x = 1 - (1-y)^{1/\beta}$ is a random variable from the beta distribution if y is a uniform random variable and β is chosen to obtain the desired expected value. We chose this form because of its simplicity and ease of calculation. The fundamental generality of species i is measured by r_i . The number of species falling within r_i measures realized generality. Occasionally, model-generated webs contain completely disconnected species or trophically identical species. Such species are eliminated and replaced until the web is free of such species. The species with the smallest n_i has $r_i = 0$ so that every web has at least one basal species.

Table 1 Basic properties of empirical food webs

Name	Taxa	S	L/S	$C(L/S^2)$
Skipwith Pond	35	25	7.9	0.32
Little Rock Lake	181	92	10.8	0.12
Bridge Brook Lake	75	25	4.3	0.17
Chesapeake Bay	33	31	2.2	0.072
Ythan Estuary	92	78	4.8	0.061
Coachella Valley	30	29	9.0	0.31
St Martin Island	44	42	4.9	0.12

'Taxa' refers to the original names for groups of organisms found in the primary reference. S refers to trophic species³. The seven food webs address (1) primarily invertebrates in Skipwith Pond¹⁵; (2) pelagic and benthic species in Little Rock Lake¹⁷, the largest food web in the primary literature; (3) Bridge Brook Lake, the largest among a recent set of 50 Adirondak lake pelagic food webs¹⁷; (4) the pelagic portion of Chesapeake Bay emphasizing larger fishes²⁰; (5) mostly birds and fishes among invertebrates and primary producers in the Ythan Estuary¹⁶; (6) a wide range of highly aggregated taxa in the Coachella desert¹⁵; and (7) trophic interactions emphasizing *Anolis* lizards on the Caribbean island of St Martin¹⁸.

(vi) Trophic similarity of a pair of species (s_{ij}) is the number of predators and prey shared in common divided by the pair's total number of predators and prey^{17,19}. We average all species' largest similarity index to calculate mean maximum similarity ($MxSim$) of a web:

$$MxSim = \frac{1}{S} \sum_{i=1}^S \max_{i \neq j} s_{ij}$$

(vii–ix) A food chain is a linked path from a species to a basal species¹⁷. The mean ($ChnLg$) and s.d. ($ChnSD$) of food chain lengths and the log of the number of food chains ($ChnNo$) are measured. Computational considerations require that chains with loops be ignored¹⁷.

(x, xi) The fraction of species that are cannibals ($Cannib$) and the fraction of species involved in longer 'loops' ($Loop$), which are food chains that include the same species twice^{5,17}.

(xii) Omnivory⁵ is the fraction of species that consume two or more species and have food chains of different lengths ($Omniv$).

Raw error is the difference between empirical properties and a model's mean predicted by Monte Carlo simulations (see Methods). We normalize raw errors by dividing them by the s.d. of the property's simulated distribution. As expected, an average of 95.8% (s.d. = 1.5, $n = 202$) of synthetic webs have properties

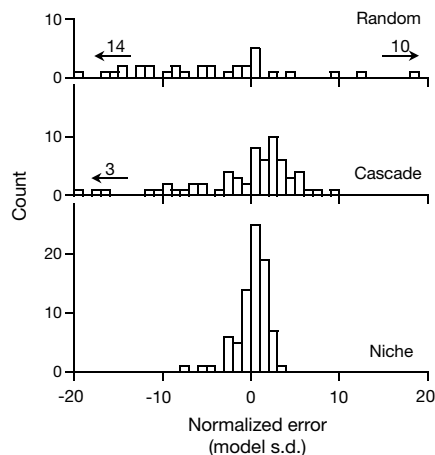


Figure 2 Distribution of normalized errors between empirical data and model means for all properties of the random, cascade and niche models. Arrows show the number of errors beyond the x-axis. Of the 56 random-model means (8 properties of 7 webs), 16% are within 2 model s.d. of the empirical data. Of the 66 cascade-model means (10 properties of 6 webs and 6 properties of one web), 27% are within this range. In contrast, 79% of 80 niche-model means (12 properties of 6 webs and 8 properties of one web) are within 2 model s.d. of the empirical data. Although attention to normalized-error magnitudes tends to reward models for increased variability, this tendency is kept in check by normalized-error s.d. < 1 that indicates excessive variability.

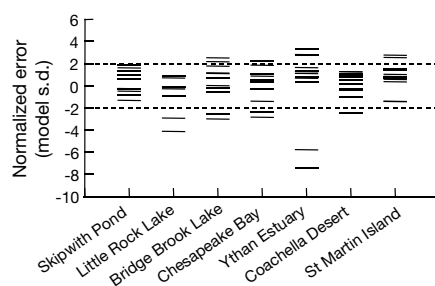


Figure 3 The niche model's normalized errors for each property of each food web. Errors are < 2 model s.d. for all properties of the Skipwith Pond web and most properties of the other webs.

within 2 model s.d. of the model's mean, which makes normalized errors between -2 and 2 a good fit because they are within the model's expected range.

Figure 2 shows the overall performance of the three models. Generally, the niche model estimates the central tendency of the empirical data remarkably well. The average normalized error is 0.22, although the s.d. of 1.8 (expected value 1) illustrates greater variability in the empirical data than in the niche model²¹ despite three distinct stochastic model components. The cascade model is over an order of magnitude worse, with an average normalized error of -3.0 and s.d. of 14.1. The random model's average of 27.1 and s.d. of 202 indicates an even worse fit. The niche model performs similarly across the different webs (Fig. 3) and consistently predicts individual properties across the group of webs more accurately than the other models (Fig. 4).

The random model's large errors show that simply matching an empirical web's S and C does little to account for empirical food-web properties except $Cannib$, which is surprisingly close to our null expectation. The cascade model improves over the random model for all properties except $Cannib$ and closely estimates T , I and B , as suggested earlier³ but previously untested against all seven webs. It

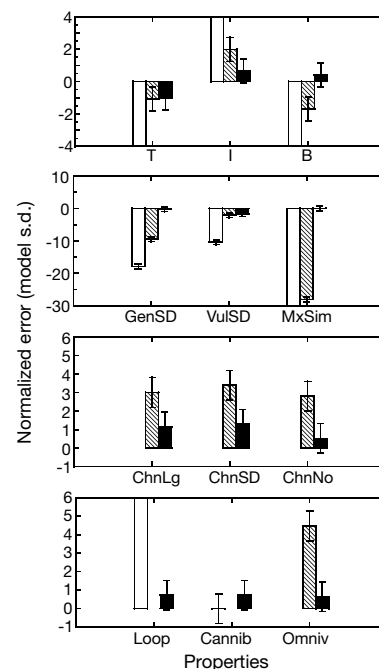


Figure 4 Mean normalized error of each property for each model averaged across the seven food webs (Table 1). The three models are indicated by open bars (random model), hatched bars (cascade model) and filled bars (niche model). Properties are described in the text. Ideally, the across-web sample average should not significantly differ from the model average of zero. Significant positive and negative average errors indicate that on average the model over- and underestimates empirical properties, respectively. Error bars show 95% confidence limits on the value of the mean assuming the empirical data are drawn from the model distribution and therefore have known population mean 0 and s.d. 1. The expected average of zero falls within the 95% confidence limits for only one property of the random model ($Cannib$), no properties of the cascade model, and eight properties of the niche model. Normalized errors do not directly correspond to raw errors because niche-model s.d. is twice as large (mean, 2.0; s.d., 0.84; $n = 66$) as cascade-model s.d. However, even in absolute terms, the magnitudes of the niche model's raw errors (Table 2) are roughly one-fifth (median 0.19, $n = 77$) of the raw errors of the random model and about one-quarter (median 0.27, $n = 80$) of the raw errors of the cascade model. In addition, the niche model has smaller average raw errors than the cascade model for all properties except T and smaller s.d. of those averages for 9 of 12 properties (see Supplementary Information). These findings show that the much greater accuracy and precision of the niche model's predictive abilities are robust to the distinction between normalized and raw errors.

also closely estimates *VulSD* but has quite large errors for other properties. The niche model improves over the cascade model for all 12 properties including *Cannib* and *Loop*, which invariably equal zero in cascade webs. This improvement is most dramatic for *MxSim*, which determines how quickly species are initially lumped in aggregation studies^{17,19} and is poorly predicted by the cascade model¹⁹.

The niche model's most significant errors may indicate problems with the data. For example, the model's underestimation of empirical variability may well be due to methodological inconsistencies among studies²¹. Also, the niche model's small but consistent overestimation of *ChnLg* and *ChnSD* (Fig. 4, Table 2) could be reduced by overcoming the well known bias against including parasites in food webs^{2,22,23}. Underestimating *T* in the Ythan Estuary web by 5.8 model s.d. (Table 2) appears to result from the web's bias towards many 'top' bird species whose consumers were excluded²². Consequently, *I* and *B* are overestimated by 3.4 and 2.8 model s.d., respectively. The other 18 empirical observations of *T*, *I* and *B* are within 1.9 model s.d. except for two at -2.3 and -2.5. Bias towards top species with zero vulnerability inflates *VulSD* and explains the niche model's underestimation of the Ythan web's *VulSD* by 7.4 model s.d.

By definition, constraining all consumers to eat one contiguous interval within a fixed sequence of species causes the niche model to generate 'interval' webs²⁰. However, larger empirical webs are rarely interval³. This discrepancy may be due to the delicacy of the intervality property. Among niche model webs with the same *S* and *C* as the seven empirical webs, intervality is broken by losing any one of almost half (mean, 41%) of the links in the webs. This suggests that we should devise a measure of the degree of intervality rather than considering intervality solely as a yes or no condition. We hypothesize that this degree is very high in empirical food webs.

A classic formulation of niche space²⁴ is of an '*n*-dimensional hyperspace' with *n* corresponding to innumerable ecological or environmental characteristics. An often-considered space that inspired our model is a species' feeding niche that restricts feeding to resources whose characteristics fall within a contiguous region of niche space^{20,24–26}. Our results show that, with respect to food-web structure, community niche space is usefully collapsible to one dimension²⁰. Whereas niche theory often infers repulsion of overlapping niches owing to interspecies competition^{24,26}, our model lacks such repulsion. Adding it or other modifications might improve the model's fit. The success for a model as simple as ours is very unexpected given the wide variety of aquatic and terrestrial food webs examined and the recently recognized complexity of their structure^{5,7–10,14–19,21,25}. The niche model merits further testing against webs from other habitats that avoid biases such as those in the Ythan web.

As it stands, the niche dimension is an empirically successful model component that facilitates a relaxed hierarchy of trophic interactions among species ordered in one dimension. Future exploration may determine the dimension's meaning and measur-

ability in the field. Such work should focus on the relationship between the mechanics of the niche model and one or more physical, evolutionary, behavioural or other mechanisms responsible for species' trophic activities. Mechanisms related to body size should be explored^{27,28}, as should algorithms that order species in empirical food webs in their 'most interval' sequence. If applying such algorithms yields strongly phylogenetic orders, evolutionary mechanisms as opposed to more conventionally invoked ecological dynamics would be suggested^{8,9,14}. Ordering algorithms would also help test the niche model's predictions, including: (1) empirical webs are close to interval; (2) species with similar niche values tend to share more predators than prey because close proximity on the niche dimension greatly increases the probability of being eaten by the same consumers while still allowing substantial differences in diet; and (3) species' niche values positively correlate with generality because species' niche ranges are products of these values (Fig. 1).

The niche model provides a benchmark for evaluating food webs as well as a structural framework to extend studies of link-strength distributions to systems larger than those previously examined^{11–13}. Link strength may be highest and lowest, respectively, for prey species near the centre and ends of a species' niche range. Although our model lacks this and many other biological mechanisms, its empirical success indicates that exploring more of the model's predictions is warranted. For example, the general effects of losing functionally distinct species on ecological systems with different levels of *S* and *C* could be predicted by simulating species losses and observing how many other species lose all their resource or consumer species. Such observations could predict extirpations due to starvation and population increases due to predation release. Effects due to species' functional traits such as omnivory could be distinguished from effects more generally due to the number of species²⁹ by simulating all possible combinations of fixed numbers of species lost. Such analyses could greatly advance scientific understanding of the potentially catastrophic consequences of species loss for the complex ecological systems on which all organisms depend. □

Methods

Monte Carlo simulations generated 1,000 webs with the same *S* and within 3% of the same *C* as an empirical web. Three per cent represents a compromise between closely matching the *C* of the empirical web and inefficiently rejecting too many model webs to find one with the empirical *C*. Several properties of some webs could not be normalized or computed. The cascade model prohibits looping and cannibalism resulting in model s.d. = 0 and raw errors that cannot be normalized. When normalized errors are discussed, these properties of the cascade model are excluded. In many random webs, *B* = 0 eliminates meaningful food-chain and omnivory properties. High *ChnLg* and *Loop* in random webs with *B* > 0 make their computation impracticable. Little Rock Lake has too many chains to compute *Omniv* or food-chain properties in a reasonable length of time for any of the models¹⁷.

Received 4 October 1999; accepted 4 January 2000.

1. May, R. M. How many species are there on earth? *Science* **241**, 1441–1449 (1988).
2. Lawton, J. H. in *Ecological Concepts* (ed. Cherratt, J. M.) 43–78 (Blackwell Scientific, Oxford, 1989).
3. Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* (Biomathematics Vol. 20) (Springer, Berlin, 1990).

Table 2 Comparison of empirically observed properties with niche-model means

Property	Skipwith Pond	Little Rock Lake	Bridge Brook Lake	Chesapeake Bay	Ythan Estuary	Coachella Desert	St Martin Island
<i>T</i>	0.040 (0.030)	0.011 (0.043)	0 (0.081)	0.32* (0.14)	0.37† (0.073)	0 (0.029)	0.17 (0.078)
<i>I</i>	0.92 (0.88)	0.86 (0.85)	0.68 (0.75)	0.52 (0.59)	0.54† (0.74)	0.90 (0.88)	0.69 (0.75)
<i>B</i>	0.040 (0.095)	0.13 (0.11)	0.32* (0.17)	0.16 (0.27)	0.090* (0.19)	0.10 (0.089)	0.14 (0.17)
<i>GenSD</i>	0.92 (0.81)	1.42† (1.08)	1.09 (1.02)	0.78* (1.13)	1.14 (1.18)	0.73 (0.82)	1.02 (1.09)
<i>VulSD</i>	0.54 (0.51)	0.61 (0.58)	0.61 (0.61)	1.12* (0.72)	1.41† (0.65)	0.60 (0.51)	0.78 (0.62)
<i>MxSim</i>	0.76 (0.75)	0.75* (0.69)	0.74* (0.62)	0.50 (0.49)	0.52* (0.59)	0.72 (0.76)	0.54* (0.62)
<i>ChnLg</i>	6.22 (7.30)		4.04 (5.75)	3.99 (4.23)	5.91 (7.37)	6.69 (7.92)	5.20 (6.73)
<i>ChnSD</i>	1.43 (1.56)		0.93* (1.48)	1.20 (1.34)	1.46 (1.87)	1.45 (1.63)	1.30 (1.67)
<i>ChnNo</i>	3.71 (4.09)		2.85 (3.10)	2.37 (2.31)	4.03 (4.36)	4.31 (4.60)	3.52 (3.86)
<i>Loop</i>	0.40 (0.69)	0.33 (0.47)	0.32 (0.27)	0 (0.030)	0 (0.098)	0.62 (0.73)	0 (0.21)
<i>Cannib</i>	0.32 (0.46)	0.14 (0.14)	0.12 (0.21)	0.032 (0.082)	0.038 (0.069)	0.66* (0.45)	0* (0.13)
<i>Omniv</i>	0.60 (0.75)		0.40* (0.60)	0.52 (0.41)	0.54 (0.60)	0.76 (0.77)	0.60 (0.62)

Niche-model means are in parentheses. See text for property descriptions. See Supplementary Information for complete results for all three models.

* Normalized errors between 2 and 3 model s.d.

† Normalized errors >3 model s.d.

4. Pimm, S. L., Lawton, J. H. & Cohen, J. E. Food web patterns and their consequences. *Nature* **350**, 669–674 (1991).
5. Polis, G. A. Complex desert food webs: an empirical critique of food web theory. *Am. Nat.* **138**, 123–155 (1991).
6. Havens, K. Scale and structure in natural food webs. *Science* **257**, 1107–1109 (1992).
7. Martinez, N. D. Effects of scale on food web structure. *Science* **260**, 242–243 (1993).
8. Martinez, N. D. Scale-dependent constraints on food-web structure. *Am. Nat.* **144**, 935–953 (1994).
9. Lawton, J. H. Webbing and WIWACS. *Oikos* **72**, 305–306 (1995).
10. Polis, G. A. & Winemiller, K. (Eds) *Food Webs: Integration of Patterns and Dynamics* (Chapman and Hall, New York, 1996).
11. deRuiter, P. C., Neutel, A.-M. & Moore, J. C. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* **269**, 1257–1260 (1995).
12. McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
13. Berlow, E. L. Strong effects of weak interactions in ecological communities. *Nature* **398**, 330–334 (1999).
14. Schoener, T. W. Food webs from the small to the large. *Ecology* **70**, 1559–1589 (1989).
15. Warren, P. H. Spatial and temporal variation in the structure of a freshwater food web. *Oikos* **55**, 299–311 (1989).
16. Hall, S. J. & Raffaelli, D. Food-web patterns: lessons from a species-rich web. *J. Anim. Ecol.* **60**, 823–842 (1991).
17. Martinez, N. D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).
18. Goldwasser, L. & Roughgarden, J. Construction of a large Caribbean food web. *Ecology* **74**, 1216–1233 (1993).
19. Solow, A. R. & Beet, A. R. On lumping species in food webs. *Ecology* **79**, 2013–2018 (1998).
20. Cohen, J. E. *Food Webs and Niche Space* (Princeton Univ. Press, Princeton, 1978).
21. Murtaugh, P. A. & Kollath, J. P. Variation of trophic fractions and connectance in food webs. *Ecology* **78**, 1382–1387 (1997).
22. Huxman, M., Raffaelli, D. & Pike, A. Parasites and food web patterns. *J. Anim. Ecol.* **64**, 168–176 (1995).
23. Marcogliese, D. J. & Cone, D. K. Food webs: A plea for parasites. *Trends Ecol. Evol.* **12**, 320–325 (1997).
24. Hutchinson, G. E. Concluding remarks. Population studies: Animal ecology and demography. *Cold Spring Harbor Symp. Quant. Biol.* **22**, 415–427 (1957).
25. Warren, P. H. in *Aspects of the Genesis and Maintenance of Biological Diversity* (eds Hochberg, M. E., Clobert, J. & Barbault, R.) 142–161 (Oxford Univ. Press, New York, 1996).
26. Begon, M., Harper, J. L. & Townsend, C. R. *Ecology: Individuals, Populations, and Communities* 3rd edn (Blackwell Science, Oxford, 1996).
27. Warren, P. H. & Lawton, J. H. Invertebrate predator–prey body size relationships: an explanation for upper triangular food webs and patterns in food web structure? *Oecologia* **74**, 231–235 (1987).
28. Cohen, J. E., Pimm, S. L., Yodanis, C. L. & Saldana, J. Body sizes of animal predators and animal prey in food webs. *J. Anim. Ecol.* **62**, 67–78 (1993).
29. Symstad, A. J., Tilman, D., Willson, J. & Knops, J. M. H. Species loss and ecosystem functioning: effects of species identity and community composition. *Oikos* **81**, 389–397 (1998).
30. Baird D. & Ulanowicz, R. E. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* **59**, 329–364 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Dunne, M. Geluardi, E. Connor, L. Goldwasser, J. Harte, T. Parker, E. Berlow and I. Billick for comments and suggestions. The USA National Science Foundation provided support.

Correspondence and requests for materials should be addressed to N.D.M. (http://userwww.sfsu.edu/~webhead; e-mail: neo@sfsu.edu).

Inhibitory threshold for critical-period activation in primary visual cortex

Michela Fagioli & Takao K. Hensch

Laboratory for Neuronal Circuit Development, Brain Science Institute RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan

Neuronal circuits across several systems display remarkable plasticity to sensory input during postnatal development^{1–10}. Experience-dependent refinements are often restricted to well-defined critical periods in early life, but how these are established remains mostly unknown. A representative example is the loss of responsiveness in neocortex to an eye deprived of vision^{2–6}. Here we show that the potential for plasticity is retained throughout life until an inhibitory threshold is attained. In mice of all ages

lacking an isoform of GABA (γ -aminobutyric acid) synthetic enzyme (GAD65), as well as in immature wild-type animals before the onset of their natural critical period, benzodiazepines selectively reduced a prolonged discharge phenotype to unmask plasticity. Enhancing GABA-mediated transmission early in life rendered mutant animals insensitive to monocular deprivation as adults, similar to normal wild-type mice. Short-term presynaptic dynamics reflected a synaptic reorganization in GAD65 knockout mice after chronic diazepam treatment. A threshold level of inhibition within the visual cortex may thus trigger, once in life, an experience-dependent critical period for circuit consolidation, which may otherwise lie dormant.

The term 'critical period' refers to a cascade of functional and anatomical events in the brain, which ultimately consolidate synaptic connections into their final wiring patterns. Once activated within visual cortex, this machinery bestows a transient sensitivity to brief monocular deprivation, which is very low just after eye opening, peaks around four weeks, and rapidly declines over the next days (rats⁴; mice⁵, Fig. 1b) or weeks (cat²; monkey³; human⁶). The critical period, however, is not a simple age-dependent maturational process, but is rather a series of events itself controlled in a

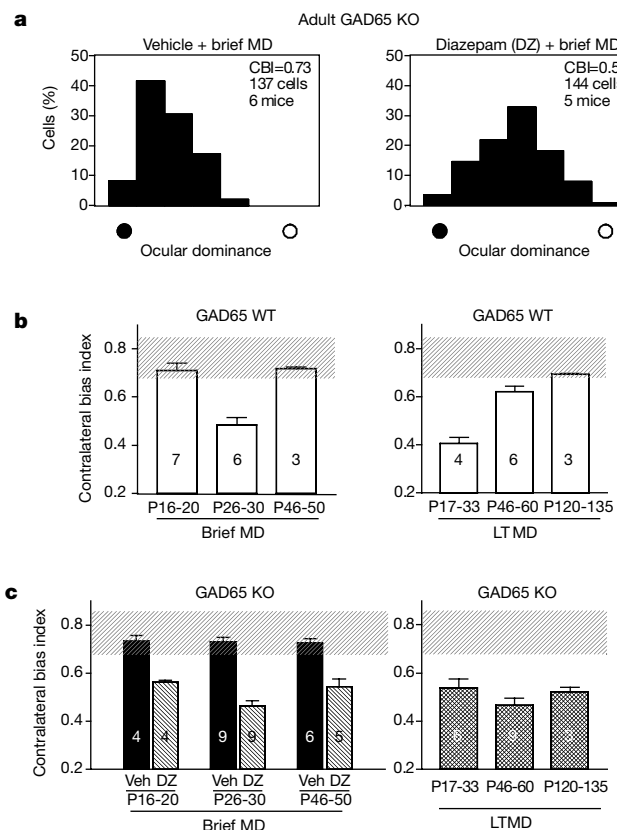


Figure 1 GAD65 knockout (KO) mice can express experience-dependent plasticity throughout life. **a**, Adult knockout mice exhibit ocular dominance shifts (diazepam versus vehicle, $P < 0.0001$, χ^2 test). **b**, Left, wild-type (WT) mice display a critical period. Brief monocular deprivation (MD) induces a significant CBI reduction only at ~P26. CBI indicates distribution bias in favour of contralateral eye (P26–30 versus non-deprived, CBI = 0.77 ± 0.02 , 6 mice, $P < 0.0001$, t -test). Right, LTMD is strongly effective only early in life. Little or no effect is detected after P45 or P120 (P17–33 versus non-deprived, $P < 0.0001$, t -test). **c**, Left, brief monocular deprivation with diazepam (DZ) infusion induces plasticity of similar strength across ages in knockout mice ($P < 0.001$ within each group, t -test). Right, LTMD is similarly effective throughout life (each group versus non-deprived, CBI = 0.77 ± 0.03 , 5 mice, $P < 0.0001$, t -test). Shaded region, range of non-deprived mice. The number of animals is indicated per group.

use-dependent manner. For example, rearing in the dark from birth delays the critical period, leaving the cortex in an immature state^{3,4} that can be altered by sensory perturbations even into adulthood^{11,12}. These results suggest that the intrinsic critical-period machinery may lie dormant until it is run down after the onset of visual experience; yet, the endogenous mechanism that triggers transient sensitivity to sensory perturbation remains unknown.

One possibility is that excitatory and inhibitory circuit elements reach an optimal balance once in life during which plasticity may occur. We have identified an animal model that does not respond to a brief monocular deprivation during the typical plastic period in wild-type mice¹³. Moreover, at this age the slight reduction of GABA release in these mice lacking the synaptic isoform of the GABA synthetic enzyme glutamic acid decarboxylase (GAD65 knockout) could be entirely compensated with benzodiazepine agonists. Such drugs enhance postsynaptic responses at active inhibitory synapses¹⁴ and restore ocular dominance plasticity¹³. We considered whether these same mice could be made to exhibit plasticity at any age, or

only during the so-called critical period. For example, a four-day period of monocular occlusion was initiated in adult GAD65 knockout mice receiving diazepam or vehicle infusion directly into one visual cortex through an osmotic minipump. Responsiveness shifted significantly in favour of the open eye only when diazepam was present during the deprivation (Fig. 1a). This plasticity did not differ from that observed previously at postnatal day, P28 (Fig. 1c; ref. 13).

This is the first example, to our knowledge, of an animal model raised with normal visual experience that retains the full potential for ocular dominance plasticity into adulthood. As in other species²⁻⁴, mice express a transient sensitivity to monocular deprivation only early in life⁵. When tested with four-day deprivations (Fig. 1b, left), plasticity is observed beginning one-week after eye-opening (P14) for about 10 days (P23–33). Long-term deprivations (LTMD, 15 days) similarly produced robust shifts in ocular dominance only when they spanned this same 10-day period (Fig. 1b, right). In contrast, GAD65 knockout mice, which do not normally respond to brief monocular deprivation on their own¹³, could be made to exhibit this plasticity before, during and after the wild-type critical period when treated with diazepam (Fig. 1c, left). Similarly, LTMD produced significant shifts in ocular dominance in GAD65 knockout animals as old as four months (Fig. 1c, right). The fact that only an extreme imbalance of input (LTMD or monocular tetrodotoxin injections as shown previously¹³) could induce plasticity without diazepam in GAD65 knockout mice strongly supports the idea that intracortical inhibition in these mutants is perpetually at a threshold for expressing plasticity.

During brief monocular deprivation, subtle pharmacological enhancement of the open probability and channel conductance of

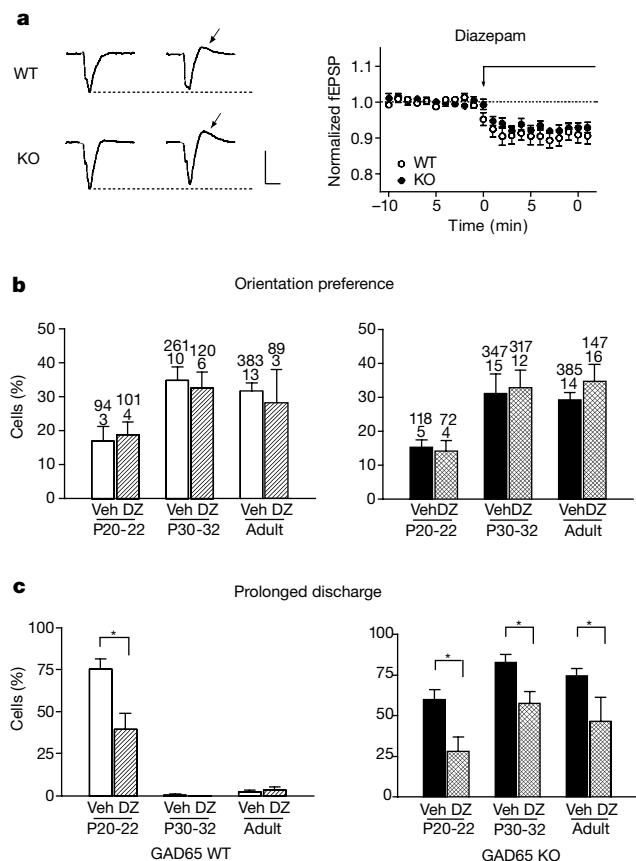


Figure 2 Diazepam reduces integrated field response and prolonged discharge. **a**, Right, acute diazepam (15 μ M, arrow) perfusion onto GAD65 wild-type (WT) or knockout (KO) cortex equally decreases peak response amplitude in layer 2/3 upon white matter stimulation (12 slices each, 8 and 6 mice, respectively). Left, representative field potential 5 min before and 10 min after drug reveal enhanced inhibitory component (arrow). Calibration, 0.5 mV, 5 ms. Responses normalized to baseline period are group data $\pm$ s.e.m. **b**, Orientation preference develops normally in both wild-type and knockout mice from P20 to adult and is not affected by chronic diazepam (DZ) treatment (P20–22 versus adult wild type (left) or adult knockout (right), both $P < 0.05$, t -test). Numbers of cells and mice are indicated above each group. **c**, Prolonged discharge is present only before critical-period onset in wild-type but throughout life in knockout mice, and is significantly decreased by diazepam in all cases (same group as in **b**; asterisk, $P < 0.05$, t -test).

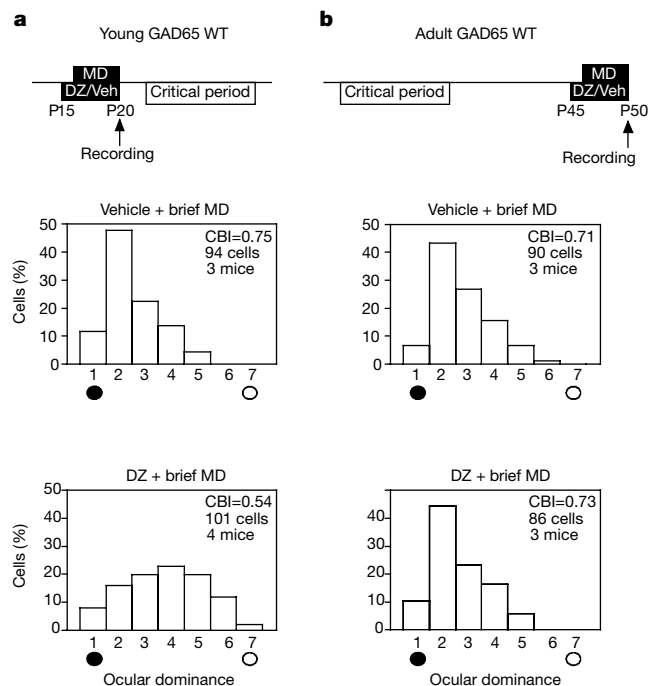


Figure 3 Enhancement of intracortical inhibition triggers plasticity in young but not adult wild-type mice. **a**, Top, young wild-type mice were treated daily (P15–20) with diazepam (DZ) or vehicle (Veh) during brief monocular deprivation (MD) initiated on P16. Brief monocular deprivation with diazepam infusion (bottom), but not vehicle (middle), shifts ocular dominance distribution in pre-critical period wild-type mice ($P < 0.0001$, χ^2 test). **b**, Top, adult wild-type mice were treated daily (P45–50) with diazepam (bottom) or vehicle (middle) during brief monocular deprivation initiated on P46. No plasticity is expressed regardless of drug treatment ($P = 0.85$, χ^2 test).

the GABA_A receptor with benzodiazepine agonists¹⁴ is sufficient to express ocular dominance plasticity. Indeed, saturating doses of diazepam only modestly decrease the integrated local field potential response in both GAD65 knockout and wild-type mice to a similar extent (Fig. 2a). Basic receptive field properties, such as size, retinotopy, response strength, habituation, spontaneous activity, orientation or direction preference were unaffected by GAD65 deletion, as reported previously¹³. We further examined the development of these features in mice to determine which might be linked to inhibitory mechanisms in accord with ocular dominance plasticity. For example, the number of cells exhibiting an orientation bias matured steadily after eye opening but was insensitive to slight alterations of inhibition by either GAD65 deletion or diazepam infusion (Fig. 2b). These properties are similarly unaffected by diazepam in kitten visual cortex (T.K.H. and M.P. Stryker, unpublished data).

Instead, a prolonged discharge of neurons upon visual stimulation, first identified at P28 in GAD65 mutants¹³, was found

throughout life in these animals (Fig. 2c, right). A similar hyperexcitability as light-bar stimuli exited the cell's receptive field was also prominent in young wild-type mice but dropped off sharply during the critical period for plasticity (Fig. 2c, left). As expected from the reduction of integrated synaptic field potentials (Fig. 2a) that underlie spike generation, prolonged discharge whenever present was significantly reduced by chronic diazepam infusion *in vivo*. Thus, ocular dominance plasticity and prolonged discharge appear to be tightly co-regulated by inhibition independent from orientation selectivity. However, we cannot rule out the possibility of subtle, higher order inhibitory effects (for example, cross-orientation), which have not yet been described in rodent visual cortex. Intracortical inhibition is generally believed to mature around the end of the first postnatal month^{15,16}. Only a particular subcircuitry may be involved, if at all^{17,18}, in determining individual receptive field properties.

If the presence of prolonged discharge indicates a potential for plasticity, then it should be possible to unmask ocular dominance plasticity in wild-type mice before the natural critical period, but not after (Fig. 2c, left). Diazepam or vehicle solution was infused one day before and throughout a four-day period of monocular eyelid suture in young (P15–20; Fig. 3a) or adult (> P45; Fig. 3b) wild-type mice. Only young animals treated with diazepam responded to monocular deprivation. Premature enhancement of inhibition, therefore, revealed the presence of a molecular machinery for plasticity which became unavailable for use in adults. These results strongly suggest that a certain level of intracortical inhibition is the endogenous trigger for activating the cascade of events underlying synapse consolidation in the critical period.

To test this idea, we attempted to restore a critical period to the GAD65 knockout mouse. We mimicked a 10-day period of sensitivity to monocular deprivation by intracranial diazepam injection, then assayed for plasticity later in life (Fig. 4a, left). In wild-type mice, LTMD after the critical period had little or no effect on ocular dominance (Fig. 4a, right; Fig. 1b, right). Similarly, adult GAD65 knockout mice that had previously been exposed to diazepam no longer exhibited plasticity. This was clearly distinct from vehicle-treated (Fig. 4b) or untreated mutants (Fig. 1c, right). The typical contralateral bias of ocular dominance was unaffected by prolonged diazepam treatment alone (contralateral bias index (CBI) = 0.72 ± 0.02 for 5 mice before, and 0.73 ± 0.01 for 3 mice after infusion). Instead, ocular dominance would have shifted had one eye been deprived concurrently, as already shown (Fig. 1c, ref. 13). Thus, a transient, early elevation of inhibition occluded future plasticity in GAD65 knockout mice, apparently re-introducing a unique window of sensitivity to sensory experience.

The specific cellular events that are triggered to actually consolidate circuit connectivity during the critical period are still unclear. We have previously shown that synaptic mechanisms of long-term potentiation (LTP) or depression (LTD) are intact in GAD65 knockout mice, despite their immunity to brief monocular deprivation¹³. Moreover, enhancing inhibitory transmission impairs LTP and LTD^{19–21}, whereas the opposite is true of experience-dependent plasticity *in vivo*¹³. Short-term synaptic dynamics have been proposed to better reflect cortical reorganization due to sensory experience²². To determine whether bursts, rather than single impulses (Fig. 2a), are processed differently in GAD65 mutants *in vitro* as *in vivo* (Fig. 2c), we recorded the synaptic response to brief trains of stimuli along vertical pathways onto layer 2/3 neurons. Short-term depression (STD) typical of these connections was significantly reduced during 5 and 10 Hz trains in GAD65 knockout mice but was restored to wild-type levels by prolonged diazepam treatment *in vivo* (Fig. 4c). Acute bath application of diazepam did not alter the magnitude of this presynaptic depression ($97 \pm 2\%$ and $98 \pm 4\%$ of pre-diazepam values at 5 Hz, 6 wild-type and 4 knockout slices, respectively). Thus, chronic restoration of inhibitory–excitatory balance to GAD65 knockout

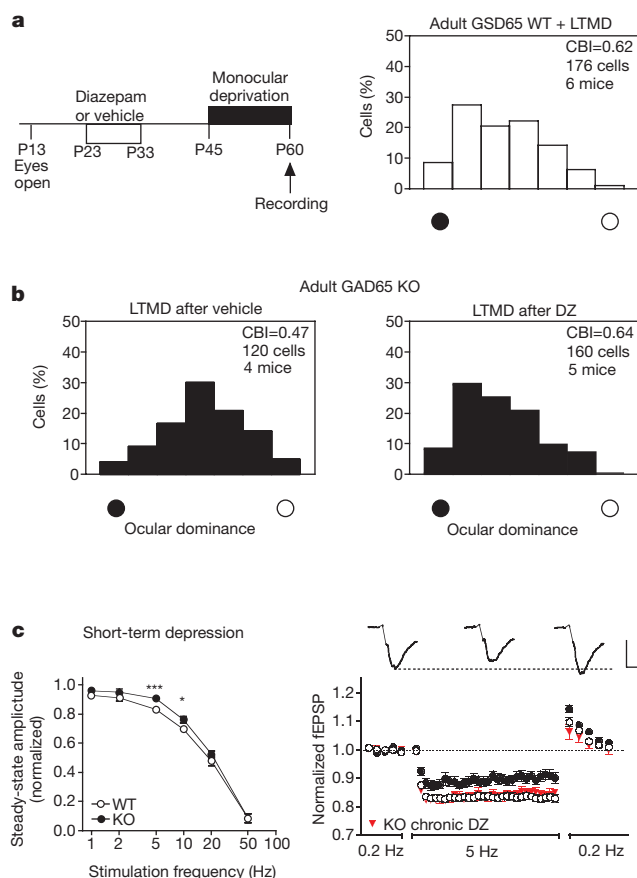


Figure 4 Early enhancement of inhibitory transmission restores a critical period and STD to GAD65 knockout mouse visual cortex. **a**, Left, knockout mice were treated daily (P23–33) with diazepam or vehicle solution before assaying ocular dominance plasticity later in life. Right, LTMD induces minimal shift of adult wild-type responses ($P < 0.05$ versus non-deprived, χ^2 test). **b**, Knockout mice treated early in life with diazepam (DZ) exhibit little plasticity, similar to adult wild-type mice ($P = 0.82$ versus wild type; $P < 0.0001$ versus knockout/vehicle, χ^2 test). **c**, Left, normalized steady-state STD for various frequency trains in GAD65 wild-type and knockout mice (6–21 slices, 4–13 animals; 7–12 slices, 5–9 animals per frequency, respectively; asterisk, $P < 0.05$; three asterisks, $P < 0.001$; t -test). Right, normalized response before, during and after a 5-Hz stimulus train with representative traces above for diazepam-treated knockout (6 slices, 4 mice; $P = 0.51$ versus wild type; $P < 0.05$ versus knockout, t -test). Plots represent mean $\pm$ s.e.m. Calibration, 0.5 mV, 2 ms.

mice may promote proper maturation of short-term synaptic dynamics, which could serve as a 'readout' of critical-period activation.

Vision is thought to initiate an irreversible developmental process of transient sensitivity to monocular deprivation^{6,12}. We have identified an animal model in which this program fails to be executed despite normal visual input. A particular inhibitory-excitatory balance within cortex may trigger the machinery for synapse consolidation, which would otherwise lie dormant. Through subtle genetic or pharmacological modulation of inhibition, it was possible to alter timing of this critical period, but it could only be initiated once in life. Thus, usage of plasticity mechanisms itself sows the seeds for terminating sensitivity to sensory perturbation, making reorganization progressively more difficult once large-scale organization of connections takes place²³. Only modest plasticity can be re-induced in adults^{24–27} by artificially tapping into processes downstream of the local circuit interactions that endogenously trigger them. Future work must elaborate which homeostatic events²⁸ are activated to set into place circuits that are no longer malleable later in life. Understanding the developmental regulation of spatio-temporal synaptic dynamics and prolonged discharge properties within visual cortex may yield deeper insight into what underlies synapse consolidation during the critical period. □

Methods

Visual cortex electrophysiology *in vivo*

Mice carrying a functional disruption of GAD65 were generated by the insertion of a neomycin-resistance cassette into exon 1 of the gene on the C57Bl/6 background, as described²⁹. Electro-physiological recordings were performed under Nembutal (50 mg kg⁻¹; Abbot)/chlorprothixene (0.2 mg; Sigma) anaesthesia using standard techniques³⁰. For each animal, 5 to 8 single units (> 75 μ m apart) were recorded in each of 4 to 6 vertical penetrations spaced evenly (> 200 μ m intervals) across the medio-lateral extent of primary visual cortex to map the monocular and binocular zones and avoid sampling bias. Cells were assigned ocular dominance scores using a 7-point classification scheme⁶. For each binocular zone, a CBI was calculated according to the formula: $CBI = [(n_1 - n_7) + (2/3)(n_2 - n_6) + (1/3)(n_3 - n_5) + N]/2N$, where N = total number of cells, and n_x = number of cells of ocular dominance score equal to x . This weighted average of the bias toward one eye or the other takes values from 0 to 1 for complete ipsilateral or contralateral eye dominance, respectively.

Monocular deprivation and drug infusion

In monocular deprivation experiments, eyelid margins were trimmed and sutured under halothane anaesthesia for 4 (brief deprivation) or 15 days (long-term deprivation). All recordings were obtained contralateral to the deprived eye and blind to drug treatment. Local rescue experiments were carried out in adult mice with low-flow osmotic minipumps (0.5 μ l h⁻¹; Alzet 1007, Alza) containing diazepam (DZ; 2mg ml⁻¹; Wako) or vehicle solution (50% propylene glycol; Wako) connected to cannulae (30 ga.) implanted stereotactically into one hemisphere under sterile surgical conditions 1 or 2 days before eyelid suture¹³. In small young mice, drug or vehicle solutions were injected daily into both lateral ventricles (1.5 μ l per hemisphere), starting one day before deprivation until the day before recording, or to control timing over a 10-day period to restore a 'critical period'. Intraventricular injections yield similar results as direct cortical minipump infusion¹³, were completed within 5 min under halothane anaesthesia, and put little stress upon the animals.

Visual cortex electrophysiology *in vitro*

Coronal slices (350 μ m thick) of mouse visual cortex were cut blind to genotype and incubated (32–33 °C) in equilibrated (95% O₂/5% CO₂, pH 7.2) artificial cerebrospinal fluid (ACSF) as described³⁰. The ACSF contained (in millimolar): 119 NaCl, 2.5 KCl, 1.3 MgSO₄, 10 NaH₂PO₄, 26.2 NaHCO₃, 2.5 CaCl₂ and 11 glucose. Slices were transferred to a recording chamber superfused with the same ACSF and maintained at 32–33 °C. Half-maximal field potential amplitudes evoked by a bipolar glass stimulating electrode filled with ACSF were recorded through a glass electrode filled with 1M NaCl (1–3 M Ω) inserted into layer 2/3 of the binocular zone. To measure diazepam sensitivity, baseline responses to white matter stimulation were recorded at a rate of 0.1 Hz for at least 15 min before applying a saturating dose of diazepam (15 μ M; Wako) to the bath.

Short-term plasticity experiments

To assay short-term presynaptic dynamics, stable, baseline responses to layer 4 stimulation were recorded at a rate of 0.2 Hz for at least 5 min, then trains (30 pulses) of various frequencies were applied with more than 1-min rest interval between each. Up to six

frequencies were tested randomly during an experiment, and each train was repeated five times. Baseline responses were carefully monitored throughout to ensure that no long-term changes in synaptic efficacy occurred. At the end of each experiment, non-NMDA and NMDA glutamate receptor antagonists, 10 μ M CNQX (Tocris) and 50 μ M D-APV (Tocris), respectively, were applied to the bath to confirm the synaptic component of the extracellular response. For analysis, maximum negative field potential amplitude was measured from the synaptic component of the raw response. Five responses before, all responses during, and five responses immediately after each stimulus train were measured and normalized to the baseline preceding each train. Steady-state depression was taken as the mean response evoked by the last 5 stimuli of the 30-pulse train. Augmentation after the completion of each train was routinely observed at the ages studied (P28–33). Chronic diazepam-treated mice received intraventricular injections as described above (> 6 days) before slicing. In acute diazepam (15 μ M) experiments, short-term plasticity was first measured in control ACSF then again after switching to perfusion medium containing the drug.

Received 8 November 1999; accepted 12 January 2000.

1. Scott, J. P. Critical periods in behavioral development. *Science* **138**, 949–958 (1962).
2. Hubel, D. H. & Wiesel, T. N. The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J. Physiol.* **206**, 419–436 (1970).
3. Daw, N. W. *Visual Development* (Plenum, New York, 1995).
4. Fagioli, M. et al. Functional postnatal development of the rat primary visual cortex and the role of visual experience: dark rearing and monocular deprivation. *Vision Res.* **34**, 709–720 (1994).
5. Gordon, J. A. & Stryker, M. P. Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *J. Neurosci.* **16**, 3274–3286 (1996).
6. Maurer, D., Lewis, T. L., Brent, H. P. & Levin, A. V. Rapid improvement in the acuity of infants after visual input. *Science* **286**, 108–110 (1999).
7. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanism. *Annu. Rev. Neurosci.* **2**, 567–631 (1999).
8. Feldman, D. E. & Knudsen, E. I. Experience-dependent plasticity and the maturation of glutamatergic synapses. *Neuron* **20**, 1067–1071 (1998).
9. Fox, K. A critical period for experience-dependent synaptic plasticity in rat barrel cortex. *J. Neurosci.* **12**, 1826–1838 (1992).
10. Crepel, F. Regression of functional synapses in the immature mammalian cerebellum. *Trends Neurosci.* **8**, 266–269 (1982).
11. Cynader, M. & Mitchell, D. E. Prolonged sensitivity to monocular deprivation in dark-reared cats. *J. Neurophysiol.* **43**, 1026–1040 (1980).
12. Mower, G. D. & Christen, W. G. Role of visual experience in activating critical period in cat visual cortex. *J. Neurophysiol.* **53**, 572–589 (1985).
13. Hensch, T. K. et al. Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science* **282**, 1504–1508 (1998).
14. MacDonald, R. L. & Olsen, R. W. GABA_A receptor channels. *Annu. Rev. Neurosci.* **17**, 569–602 (1994).
15. Blue, M. E. & Parnevelas, P. J. The formation and maturation of synapses in the visual cortex of the rat. II. Quantitative analysis. *J. Neurocytol.* **12**, 697–712 (1983).
16. Luhmann, H. J. & Prince, D. A. Postnatal maturation of the GABAergic system in rat neocortex. *J. Neurophysiol.* **65**, 247–263 (1991).
17. Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
18. Nelson, S., Toth, L., Sheth, B. & Sur, M. Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science* **265**, 774–777 (1994).
19. Bear, M. F. & Kirkwood, A. Neocortical long-term potentiation. *Curr. Opin. Neurobiol.* **3**, 197–202 (1993).
20. Malinow, R. & Miller, J. P. Postsynaptic hyperpolarization during conditioning reversibly blocks induction of long-term potentiation. *Nature* **320**, 529–530 (1986).
21. del Cerro, S., Jung, M. & Lynch, G. Benzodiazepines block long-term potentiation in slices of hippocampus and pyriform cortex. *Neuroscience* **49**, 1–6 (1992).
22. Finnerty, G. T., Roberts, L. S. E. & Connors, B. W. Sensory experience modifies the short-term dynamics of neocortical synapses. *Nature* **400**, 367–371 (1999).
23. Knudsen, E. I. Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science* **279**, 1531–1533 (1998).
24. Pettigrew, J. D. & Kasamatsu, T. Local perfusion of noradrenaline maintains visual cortical plasticity. *Nature* **271**, 761–763 (1978).
25. Mataga, N., Imamura, K. & Watanabe, Y. L-threo-3,4-dihydroxyphenylserine enhanced ocular dominance plasticity in adult cats. *Neurosci. Lett.* **142**, 115–118 (1992).
26. Muller, C. M. & Best, J. Ocular dominance plasticity in adult cat visual cortex after transplantation of cultured astrocytes. *Nature* **342**, 427–430 (1989).
27. Gu, Q., Liu, Y. & Cynader, M. S. Nerve growth factor induced ocular dominance plasticity in adult cortex. *Proc. Natl Acad. Sci. USA* **91**, 8408–8412 (1994).
28. Turrigiano, G. G. Homeostatic plasticity in neuronal networks: the more things change the more they stay the same. *Trends Neurosci.* **22**, 221–227 (1999).
29. Kash, S. F. et al. Epilepsy in mice deficient in the 65-kDa isoform of glutamic acid decarboxylase. *Proc. Natl Acad. Sci. USA* **94**, 14060–14065 (1997).
30. Hensch, T. K. et al. Comparison of plasticity *in vivo* and *in vitro* in the developing visual cortex of normal and protein kinase A R1B-deficient mice. *J. Neurosci.* **18**, 2108–2117 (1998).

Acknowledgements

We thank K. Hartman for contributing to Fig. 4c, N. Mataga for comments, and S. Fujishima for genotyping and maintenance of the GAD65 mouse colony re-derived from original heterozygote breeding pairs which were kindly provided by S. Baekkeskov and S.F. Kash.

Correspondence and requests for materials should be addressed to T.K.H. (e-mail: hensh@postman.riken.go.jp).

Attention modulates synchronized neuronal firing in primate somatosensory cortex

P. N. Steinmetz⁺, A. Roy, P. J. Fitzgerald, S. S. Hsiao, K. O. Johnson & E. Niebur

Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, Maryland 21218, USA

⁺Present address: Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, USA

A potentially powerful information processing strategy in the brain is to take advantage of the temporal structure of neuronal spike trains. An increase in synchrony within the neural representation of an object or location increases the efficacy of that neural representation at the next synaptic stage in the brain; thus, increasing synchrony is a candidate for the neural correlate of attentional selection¹. We investigated the synchronous firing of pairs of neurons in the secondary somatosensory cortex (SII) of three monkeys trained to switch attention between a visual task and a tactile discrimination task. We found that most neuron pairs in SII cortex fired synchronously and, furthermore, that the degree of synchrony was affected by the monkey's attentional state. In the monkey performing the most difficult task, 35% of neuron pairs that fired synchronously changed their degree of synchrony when the monkey switched attention between the tactile and visual tasks. Synchrony increased in 80% and decreased in 20% of neuron pairs affected by attention.

Each monkey was trained to perform tactile and visual tasks and to switch between them when cued. The visual task was a dimming detection task: three white squares appeared on a computer screen and after a random interval one of the squares, selected at random, dimmed slightly. The tactile stimuli continued unabated during the visual task and were not coincident with the visual stimuli. Each monkey performed a different tactile task. Two monkeys discriminated raised letters (6.0 mm high) scanned across a distal fingerpad (15 mm s⁻¹), pressing a key when the letter on the finger matched the target letter displayed on a computer screen². The height of the tactile letters was chosen to be close to the resolution limit in humans; the monkeys' performance was identical to that in humans discriminating the same letters². The target letter displayed on the computer screen was large (>3° high) and stayed on continuously during the tactile task. For monkey M1, the target letter remained constant within trials in which a single set of neurons was studied (~45 min). For monkey M2, the target letter changed randomly after each correct response (every third or fourth letter on average; that is, about every 7.5–10 s). Monkey M3 discriminated whether bars (6.0 mm long) presented successively to a distal fingerpad had the same or different (by 90°) orientations. All three tactile tasks are difficult for humans, but M2's task is particularly taxing because of the continually changing tactile targets. The monkeys' responses were about 90% correct in all tasks. Each monkey was cued to switch between the tactile and visual tasks about every 7–8 min while simultaneous single-unit recordings were made from up to seven microelectrodes³ located in the contralateral SII cortex an area known to be affected by attention^{2,4,5}.

Spike train data from each neuron pair were first sorted into blocks according to stimulus (particular letter or bar orientation) and task (tactile or visual). Two cross-correlograms were computed for each block—a raw correlogram between simultaneously

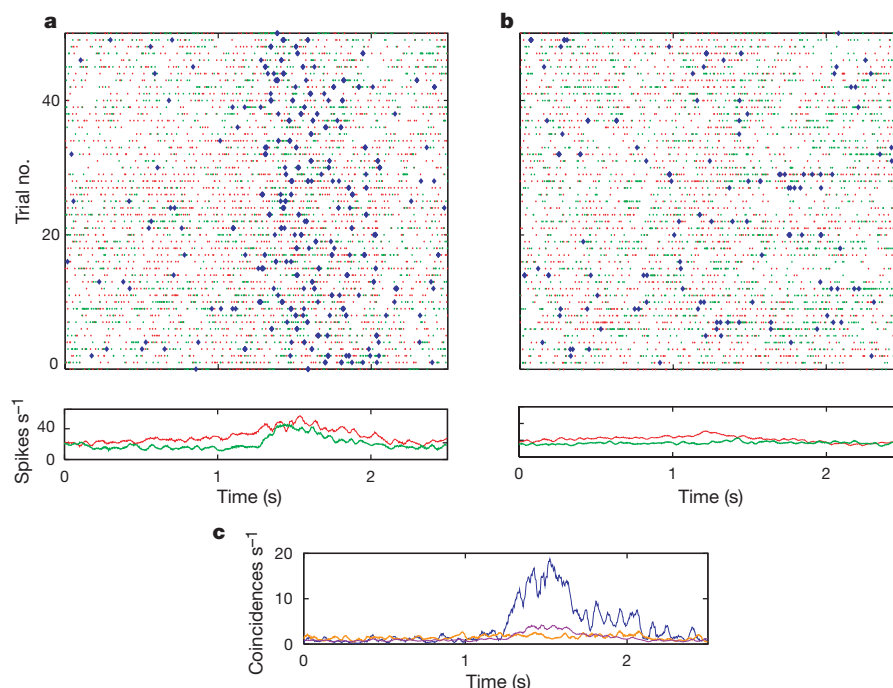


Figure 1 Responses of a typical neuron pair in monkey M2. Responses are triggered, at the onsets of 50 tactile stimulus periods, while the monkey is performing the tactile task (a) and the visual task (b). Each row in the raster represents one stimulus period, 2.5 s long, corresponding to the presentation of one letter. The letter enters the receptive field at about time (t) = 1.25 s. Red and green dots represent the action potentials of the neuron pair. Peristimulus time histograms (smoothed with a gaussian filter, s.d. 10 ms) are shown below each raster plot with corresponding colours. Synchronous events, defined as spikes

from each neuron within 2.5 ms of each other, are represented as blue diamonds. The number of synchronous events is much higher when attention is directed towards the tactile stimuli. This change in synchrony is also apparent in the smoothed plots of rates of synchronous events shown in c; blue curve: tactile task; orange curve: visual task; violet curve: rate of synchrony expected by chance in tactile task. Note that all synchronous events are shown here whereas the statistical analyses subtract the estimated profiles of chance synchronous events.

recorded pairs of responses and a 'shift-predictor' correlogram between all nonsimultaneous pairs, which estimates the synchrony expected by chance. The shift-predictor corrected cross-correlogram (SCCC), which measures synchrony above or below that expected by chance, was computed by subtracting the shift-predictor correlogram from the raw correlogram. Synchrony was analysed in a 50-ms window (± 25 ms around zero delay). Fifty milliseconds was chosen because about 70% of the neuron pairs with significant correlation peaks had half widths less than 50 ms, and 50 ms is approximately the period of perceptual integration for tactile stimuli⁶.

Changes in synchrony between the visual and tactile tasks were assessed by computing the sum of squared differences between the bins of the visual and tactile SCCC's. We analysed only neuron pairs in which the spikes were collected on separate electrodes (400 μ m minimal spacing) to ensure that the neurons were distinct. Spikes were determined visually to be well isolated from background noise and other spikes in all cases. The mean distance between recording sites exceeded 1,000 μ m.

We analysed 648 pairs of responses from 436 neurons in SII cortex in four hemispheres of three monkeys. Seventy-eight per cent (339/436) of these neurons showed a significant change in firing rate

when the animal switched between the tactile and visual tasks². Sixty-six per cent (427/648) of the neuron pairs had significant cross-correlogram peaks ($P < 0.05$) during the visual task, the tactile task or both (Table 1). The high percentage of neuron pairs with correlated firing may be accounted for by the large, overlapping receptive fields in SII cortex.

Seventeen per cent (74/427) of the neuron pairs with significant cross-correlogram peaks also showed significant changes in synchrony between the visual and tactile tasks. Eighty per cent (59/74) responded with increased and 20% responded with decreased synchrony when the monkey performed the tactile task. Statistically significant synchrony was observed during both the visual and tactile tasks in most of these pairs. Figure 1 shows raster plots for a neuron pair in SII cortex of monkey M2 in which attention to the tactile stimulus produced a significant increase in synchrony. Synchronous events (spikes from two neurons within 2.5 ms of each other) are represented as large blue diamonds. The rate of synchronous events (Fig. 1c) rises to almost 20 s^{-1} as the stimulus letter passes over the fingerpad in the tactile task. During the visual task, the rate rarely exceeds 2 s^{-1} . A certain rise in synchrony is expected because the impulse rates in both neurons rise during the tactile task (Fig. 1a) but the rate of synchronous events expected by chance never rises above 4 s^{-1} (Fig. 1c, violet curve).

Representative SCCC's from each of the three monkeys are shown in Fig. 2. Figure 2a and b illustrates pairs with relative small and moderate increases in synchrony, respectively, when the animal switched to the tactile task. Figure 2c illustrates a neuron pair in which synchrony decreased. The abscissa on each inset histogram represents the mean change in the coincidence rate over the entire trial duration, the change in coincidence rate during the response is much larger than the mean value displayed on the abscissa.

The fractions of all neuron pairs with significant synchrony that also had significant changes in synchrony ($P < 0.05$) between the tactile and visual tasks (Table 1) were 16.0, 35.3 and 9.5% for M1, M2 and M3, respectively. When the criterion for significance was set at $P < 0.01$ the respective values were 8.6, 28.0 and 6.6%. When stated in terms of neurons rather than pairs the percentages were larger: of all the neurons involved in synchronous firing, 18.5, 53 and 26% in M1, M2 and M3, respectively were involved in one or more pairings in which synchrony changed significantly ($P < 0.05$). The probability of obtaining any of these percentages by chance is 0.003 in the worst case (M1 tested at $P < 0.05$; that is, the probability of obtaining 8 (16%) or more significant results in 50 cases when chance alone yields 2.5 cases (5%) is 0.003).

It is not surprising that only a fraction of the neurons in the population show significant changes in synchrony if change in synchrony is an important attentional mechanism. Neurons play different roles in perceptual processing. If only a subset of the neurons in a region are engaged in representing the attended stimulus, then only that subset should show increased synchrony. It is also not surprising that the fraction of neurons showing changes in synchrony is larger than the fraction of pairs; for example, if 30% of neurons belong to a particular subset then only 9% of all possible pairs draw both members from that subset.

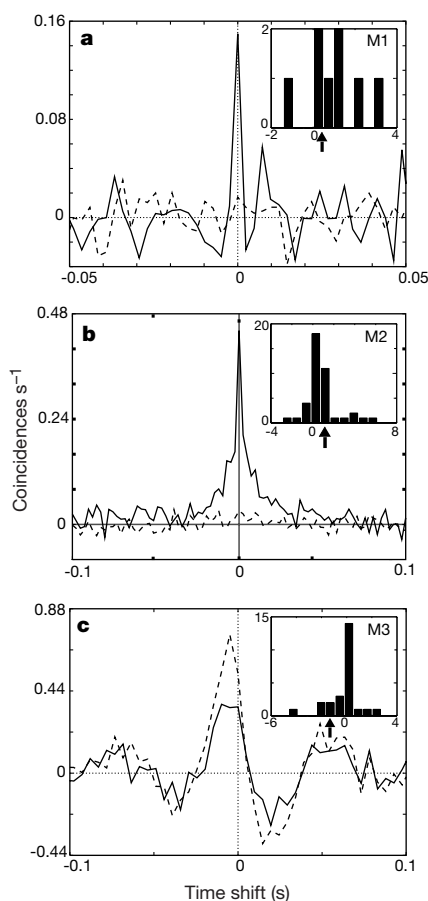


Figure 2 Shift predictor corrected cross-correlograms (SCCCs). Plots are for three pairs of neurons from monkeys M1 (a), M2 (b) and M3 (c). SCCC for the tactile and visual tasks are represented by the solid and dashed lines, respectively. The ordinate of each SCCC displayed here has been normalized to coincidences per second to account for differences in trial duration by dividing the counts in each SCCC bin by trial duration. The sum of differences between normalized tactile and visual SCCC bins (over ± 25 ms) yields an overall measure of the change in rate of synchronous events. Insets show histograms of this overall change in rate of synchrony (coincidences s^{-1}) for all pairs with a significant attentional effect (see Table 1). Negative numbers in the histogram correspond to decreased synchrony in the tactile task. The arrow below each histogram identifies the change in rate of synchronous firing for the SCCC pair shown.

Table 1 Degree of synchrony in SII cortex

Monkey*	Synchrony†	Change‡	Increase§
M1	50/95 (52.6%)	8/50 (16.0%)	7/8 (87.5%)
M2	113/145 (77.9%)	41/116 (35.3%)	35/41 (85.4%)
M3	264/408 (64.7%)	25/264 (9.5%)	17/25 (68.0%)
Average	427/648 (65.9%)	74/427 (17.3%)	59/74 (79.7%)

*M1: letter discrimination, constant target; M2: letter discrimination, varied target; M3: bar orientation discrimination.

† Fraction of cell pairs with significant synchronous firing ($P < 0.05$).

‡ Fraction of neuron pairs in column two in which the synchrony changed between the visual and tactile tasks ($P < 0.05$).

§ Percentage of neuron pairs in column three in which synchrony increased during the tactile task.

The differences in numbers of affected neuron pairs between monkeys may have more than one explanation. One possibility is that the number of neurons engaged in representing statically indented bars is smaller than the number engaged by scanned complex forms such as letters; however, this does not explain the substantial difference between M1 and M2, both of which performed letter discrimination tasks. A second possibility, which may explain the difference between M1 and M2, is that the attentional demands differed between tasks. Human reaction times and error rates in a discrimination task are lower when the same reference pattern is presented repeatedly than when the reference pattern changes on every trial^{7,8}. This is the principal difference between the tasks performed by M1 and M2. There is extensive evidence that the cognitive loads in the two experimental designs are different and that the second design is more taxing^{7,8}. This difference in cognitive load may account for the larger percentage of neurons affected by attention in M2.

The temporal structure of spike trains has been suggested previously as a basis for information processing in the brain^{9,10} and there are numerous reports of synchronous firing in the primate central nervous system^{11–20}. Here, we investigated whether the degree of synchrony is modifiable and whether it changes when the animal changes its attentional focus. The changes in synchrony cannot be explained by stimulus-driven events as the tactile stimulus was identical in the visual and tactile tasks; nor can it be explained by attention-induced modulation of the neuronal firing rates, because we controlled for this effect with a shift predictor. The latter point was confirmed by an analysis showing that changes in the rate of synchrony were uncorrelated with changes in firing rate (measured as the change in summed rate; correlation, -0.0243). A lack of correlation between impulse rate and synchrony has been reported in a different context^{14–16}.

The changes in synchrony reported here are consistent with those predicted by theoretical analyses²¹ and a computational model of attention¹. That model, which motivated our study, accounts for the selectivity inherent in attention by a mechanism that increases synchrony in the representation of the location being attended in order to increase its synaptic efficacy. Whether that model is correct or not, increasing (decreasing) synchrony between neurons is a powerful mechanism for increasing (decreasing) the combined synaptic effect of a subset of neurons. If a significant number of the neurons that cooperate in the distributed representation of an object or location in space fire more (less) synchronously than neurons representing other objects or locations, then their combined message is more (less) potent. Change in synchrony may be an essential neural mechanism of selective attention. □

Methods

Animals and animal training

All surgical and experimental procedures were approved by the Animal Care and Use Committee of the Johns Hopkins University. Monkeys (male, rhesus, 3–5 kg) were placed on a restricted water diet and brought into the laboratory on 6 days a week for 4–5 months for training. They twisted a manual switch (M1) with their ipsilateral hand or pushed and pulled a foot switch (M2 and M3) to signal their responses and receive liquid rewards.

Recordings

Neural activity was recorded using seven separate extracellular microelectrodes driven by a Reitboeck microdrive³. The end of the microdrive was modified so that the seven electrodes were linearly aligned and spaced 400 or 600 μm apart. Electrode tracks were marked with dyes and standard histological techniques were used to confirm the recording sites in SII cortex²². Individual spikes were isolated using both a window amplitude discriminator and a template-based discriminator (Alpha Omega Corp). Only spikes that were determined visually to be well isolated and stable over the entire recording session were included in the analysis.

Data analysis

Blocks of tactile trials (animal performing the tactile task) were alternated with blocks of visual trials. Trial duration was 2.5 s (for M1 and M2) or 4.5 s (M3) and each analysis was based on at least four alternations between blocks². A trial consisted of the presentation of

one letter in the letter-discrimination task or a pair of bars in the bar-orientation task. Only tactile trials with correct responses were included in the analyses. Excitability covariation generated by long-term changes in rate²³ is unlikely (average correlation between paired rates was 0.055). Each trial was divided into 1,024 bins which yielded bin widths of 2.4 ms (M1 and M2) or 4.4 ms (M3). Raw cross-correlograms between pairs of neuronal responses were computed for individual trials and then averaged across all trials with the same stimulus (a specific letter or specific member of an orientation pair) and task (tactile or visual); this constituted an analysis block. The correlogram expected by chance was constructed by computing individual cross-correlograms between the responses of one neuron of the pair in each trial and the responses of the other neuron in every non-corresponding trial in the same analysis block and then averaging those correlograms. This average chance correlogram (the shift predictor) was subtracted from the average raw correlogram to obtain an estimate of the synchrony above or below that expected by chance. The final averaged shift-predictor corrected cross-correlogram (SCCC) for each task was then computed by averaging over all corrected correlograms.

The sum of squared SCCC bin values was used as the summary measure of synchrony. The sum of the squared differences, on a bin-by-bin basis, between SCCC for the tactile and visual tasks was used as the measure of change in synchrony. Similar results were obtained with other test statistics (for example, sums and differences of absolute values). The first question addressed was whether individual SCCC were different from those expected from chance synchrony between independent responses. This null hypothesis was tested with the Fisher permutation test²⁴. The original data set was replicated exactly except that the response of one neuron in each trial was paired with a response of the other neuron, selected randomly and without replacement from another trial with the same stimulus and task. Then the analysis was repeated exactly as with the original data. Five hundred such replications produced the distribution of synchrony values that would have arisen if the neurons fired independently²⁵. The second question was whether synchrony differed significantly between the tactile task and the visual task (the baseline condition). The null hypothesis is different from that in the first question above; it is that tactile and visual SCCC may both display statistically significant synchrony but that they differ from one another by no more than a pair of visual SCCC drawn at random. Random samples of the visual SCCC were constructed by a bootstrap procedure²⁴, individual visual SCCC were constructed by averaging single-trial cross-correlograms drawn randomly with replacement from the visual single-trial correlograms. Differences that would occur by chance under the null hypothesis were computed as the sum of squared differences between pairs of these visual SCCC. The significance of the observed difference between the SCCC derived from the tactile and visual tasks was computed as the fraction of differences resulting from the null hypothesis that exceeded the observed difference. Virtually identical results were obtained with bootstrap tests based on tactile SCCC and a bootstrap version of the *t*-test. The use of these bootstrap procedures provides a model-free test of the degree of synchrony and does not assume any stochastic model of neural firing (for example, Poisson) or independence of firing in neighbouring time analysis bins²⁵.

Received 8 November; accepted 27 December 1999.

- Niebur, E. & Koch, C. A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J. Comput. Neurosci.* **1**, 141–158 (1994).
- Hsiao, S. S., O'Shaughnessy, D. M. & Johnson, K. O. Effects of selective attention on spatial form processing in monkey primary and secondary somatosensory cortex. *J. Neurophysiol.* **70**, 444–447 (1993).
- Mountcastle, V. B., Reitboeck, H. J., Poggio, G. F. & Steinmetz, M. A. Adaptation of the Reitboeck method of multiple microelectrode recording to the neocortex of the waking monkey. *J. Neurosci. Methods* **36**, 77–84 (1991).
- Poranen, A. & Hyvärinen, J. Effects of attention on multiunit responses to vibrations in the somatosensory regions of the monkey's brain. *EEG Clin. Neurophysiol.* **53**, 525–537 (1982).
- Burton, H., Sinclair, R. J., Hon, S.-Y. & Whang, K. C. Tactile-spatial and cross-modal attention effects in the second somatosensory and 7b cortical areas of rhesus monkey. *Somatosens. Mot. Res.* **14**, 237–267 (1997).
- Craig, J. C. Interference in identifying tactile patterns: response competition and temporal integration. *Somatosens. Mot. Res.* **13**, 188–213 (1996).
- Shiffrin, R. M. & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* **84**, 127–190 (1977).
- Schneider, W. & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search and attention. *Psychol. Rev.* **84**, 1–66 (1977).
- Poggio, G. F. & Viernstein, L. J. Time series analysis of impulse sequences of thalamic somatic sensory neurons. *J. Neurophysiol.* **27**, 517–545 (1964).
- Perkel, D. H., Gerstein, G. L. & Moore, G. P. Neuronal spike trains and stochastic point processes. II: Simultaneous spike trains. *Biophys. J.* **7**, 419–440 (1967).
- Abeles, M. *Local Cortical Circuits* (Springer, Berlin, Heidelberg, New York, 1982).
- Eckhorn, R. et al. Coherent oscillations: a mechanism of feature linking in the visual cortex? *Biol. Cybern.* **60**, 121–130 (1988).
- Abeles, M., Bergman, H., Margalit, E. & Vaadia, E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* **70**, 1629–1638 (1993).
- Vaadia, E. et al. Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518 (1995).
- Prut, Y. et al. Spatiotemporal structure of cortical activity: properties and behavioral relevance. *J. Neurophysiol.* **79**, 2857–2874 (1998).
- Haalman, I. & Vaadia, E. Emergence of spatio-temporal patterns in neuronal activity. *Z. Naturforsch.* **53C**, 657–669 (1998).
- Murthy, V. N. & Fetz, E. E. Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys. *J. Neurophysiol.* **76**, 3968–3982 (1996).
- Singer, W. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.* **55**, 349–374 (1993).

19. Decharms, R. C. & Merzenich, M. M. Primary cortical representation of sounds by the coordination of action potential timing. *Nature* **381**, 610–613 (1996).
20. Roy, S. & Calloway, K. D. Synchronization of Local Neural Networks in the Somatosensory Cortex. A Comparison of Stationary and Moving Stimuli. *J. Neurophysiol.* **81**, 999–1013 (1999).
21. Crick, F. & Koch, C. Towards a neurobiological theory of consciousness. *Sem. Neurosci.* **2**, 263–275 (1990).
22. DiCarlo, J. J., Lane, J. W., Hsiao, S. S. & Johnson, K. O. Marking microelectrode penetrations with fluorescent dyes. *J. Neurosci. Methods* **54**, 75–81 (1996).
23. Brody, C. D. Slow covariations in neuronal resting potentials can lead to artefactually fast cross-correlations in their spike trains. *J. Neurophysiol.* **80**, 3345–3351 (1998).
24. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).
25. Roy, A., Steinmetz, P. N., Johnson, K. O. & Niebur, E. Model-free detection of synchrony in neuronal spike trains, with an application to primate somatosensory cortex. *Neurocomputing* (in the press).

Acknowledgements

This work was supported by the NIH, the NSF and the Alfred P. Sloan Foundation. We thank J. DiCarlo, M. Usher and S. Yantis for discussions and J. Lane for technical support.

Correspondence and requests for materials should be addressed to E.N.
(e-mail: niebur@jhu.edu).

Growth patterns in the developing brain detected by using continuum mechanical tensor maps

Paul M. Thompson*, Jay N. Giedd†, Roger P. Woods*, David MacDonald‡, Alan C. Evans‡ & Arthur W. Toga*

*Laboratory of Neuro Imaging, Department of Neurology, Division of Brain Mapping, UCLA School of Medicine, 710 Westwood Plaza, Los Angeles, California 90095-1769, USA

†Child Psychiatry Branch, National Institute of Mental Health, NIH, 10 Center Drive, MSC 1600, Bethesda 20982-1600, Maryland, USA

‡Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Québec, Canada H3A 2B4

The dynamic nature of growth and degenerative disease processes requires the design of sensitive strategies to detect, track and quantify structural change in the brain in its full spatial and temporal complexity¹. Although volumes of brain substructures are known to change during development², detailed maps of these dynamic growth processes have been unavailable. Here we report the creation of spatially complex, four-dimensional quantitative maps of growth patterns in the developing human brain, detected using a tensor mapping strategy with greater spatial detail and sensitivity than previously obtainable. By repeatedly scanning children (aged 3–15 years) across time spans of up to four years, a rostro-caudal wave of growth was detected at the corpus callosum, a fibre system that relays information between brain hemispheres. Peak growth rates, in fibres innervating association and language cortices, were attenuated after puberty, and contrasted sharply with a severe, spatially localized loss of subcortical grey matter. Conversely, at ages 3–6 years, the fastest growth rates occurred in frontal networks that regulate the planning of new actions. Local rates, profiles, and principal directions of growth were visualized in each individual child.

Time series of high-resolution three-dimensional magnetic resonance imaging (MRI) scans were acquired across large time spans from young normal subjects (aged 3–6, 6–7, 7–11, 8–12, 9–13 and 11–15 years) at intervals ranging from two weeks to four years. Growth patterns were recovered by computing a three-dimensional elastic deformation field, which reconfigures the anatomy at the earlier time point into the shape of the anatomy of the later scan.

Maps of local growth rates (Figs 1–4) revealed the complexity and regional heterogeneity of the tissue growth, pruning and maturation processes of late brain development. In subjects aged 6–15 years, the

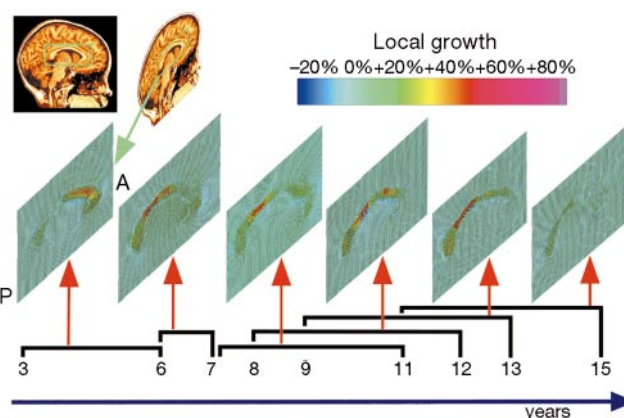


Figure 1 Growth patterns in the developing human brain detected at ages 3–15 years. A rostro-caudal wave of peak growth rates is detected in young normal subjects scanned repeatedly across time spans of up to four years. Between ages 3 and 6 years, peak growth rates (red colours; 60–80% locally) were detected in the frontal circuits of the corpus callosum, which sustain mental vigilance and regulate the planning of new actions. Older children displayed fastest growth at the callosal isthmus, which innervates temporoparietal systems supporting spatial association and language function. Between ages 11–15 years, growth rates still peak at the isthmus, but are attenuated.

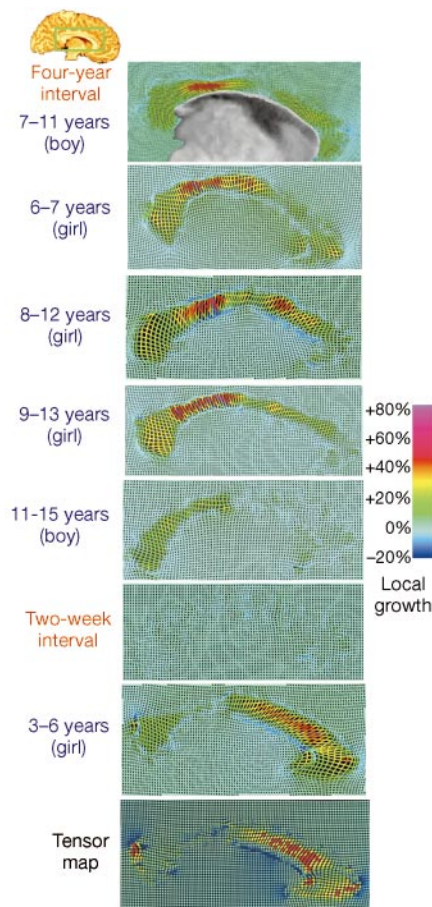


Figure 2 Mapping dynamic patterns of brain development: four-dimensional growth maps. Strikingly similar growth rates were detected in the corpus callosum of five young normal subjects scanned repeatedly aged 6–13 years. Peak values throughout the posterior midbody (red colours) were attenuated after puberty (11–15 years). By contrast, near-zero maps of change were observed between scans acquired over a two-week interval. Between ages 3–6 years, extreme growth rates were found in the anterior interhemispheric fibre systems that transfer information to sustain mental vigilance and organize new actions. Tensor maps identify the principal directions of growth rates, revealing an outward radial tissue expansion in frontal regions.

highest growth rates were consistently attained in temporo-parietal systems which are functionally specialized for language, and for understanding spatial relations (Fig. 2). In contrast to the near-zero maps of change recovered at short time intervals ('Two-week interval' in Fig. 2), growth maps spanning large time intervals showed complex and heterogeneous patterns of change. Between ages 7 and 11 years (Fig. 2), comparative stability of the splenial and rostral fibre systems of the corpus callosum contrasted sharply with

rapid focal growth at the callosal isthmus (up to 80%). Although global measurements indicated an overall 22.4% increase in mid-sagittal callosal area during the four-year time span (from 527.6 mm² to 645.6 mm²), these global values disguise the complexity of local growth patterns. Local growth is as high as 80% (Fig. 2), a feature which may not be apparent with conventional volumetric descriptors.

Although some individual variation was expected, this focus of extreme growth at the callosal isthmus was detected consistently in all subjects tracked between 6 and 15 years (Fig. 2), suggesting that cortico-cortical networks supporting rapid associative relay and language functions may myelinate more extensively³ and over longer periods than rostral fibre systems. In a girl scanned twice exactly one year apart at ages 6 and 7 years, extreme growth (up to 85%) at the callosal isthmus contrasted with a comparatively quiescent region in the more rostral systems that innervate frontal and pre-frontal cortices. When a four-year growth map was generated for a slightly older child (11–15 years, Fig. 2), growth rates were correspondingly reduced in every region. Nonetheless, growth patterns at the isthmus and splenium (commonly defined as the posterior fifth of the callosum) were still more rapid (20–25% locally) than in the more anterior rostrum and genu (near-zero change). In an analysis of grey matter at the cortex⁴, we recently observed a localized grey matter loss in frontal cortex that persists in normal subjects throughout adolescence even into adulthood. The gradual quiescence of growth at the rostral callosum around puberty may therefore be a precursor to a prolonged regressive process of grey matter loss through adolescence into adulthood in the frontal circuits it innervates.

Several near-zero maps of change were recovered at short time intervals. Figure 2 shows a typical map from a subject scanned at age 8 years, exactly four years later at age 12 years, and again two weeks later. Negligible change at short time intervals ('Two-week interval' in Fig. 2) contrasted with a highly heterogeneous map of growth across the four-year time span. Growth rates again achieved their highest rates in the associative and linguistic networks that cross at the callosal isthmus.

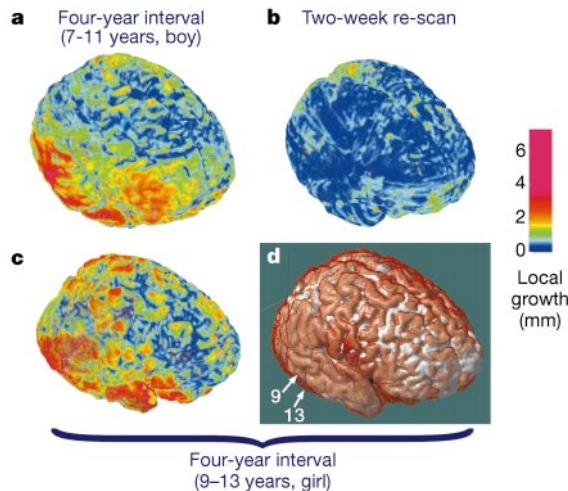


Figure 3 Patterns of cerebral growth. **a**, In a subject scanned at age 7 years and again exactly four years later at age 11 years, dramatic growth is found in temporo-parietal regions (red colours). **b**, All brain regions are stable in a control experiment (blue colours) analysing scans acquired two weeks apart. **c**, Between ages 9–13 years, growth is also most pronounced in temporo-parietal regions. **d**, This was confirmed by digitally overlaying models of the cerebral cortex at each time point (arrows 9 and 13). Growth at the callosal isthmus in subjects aged 7–11 and 9–13 years (Fig. 2) is therefore accompanied by diffuse growth in its (temporo-parietal) lobar projection zones.

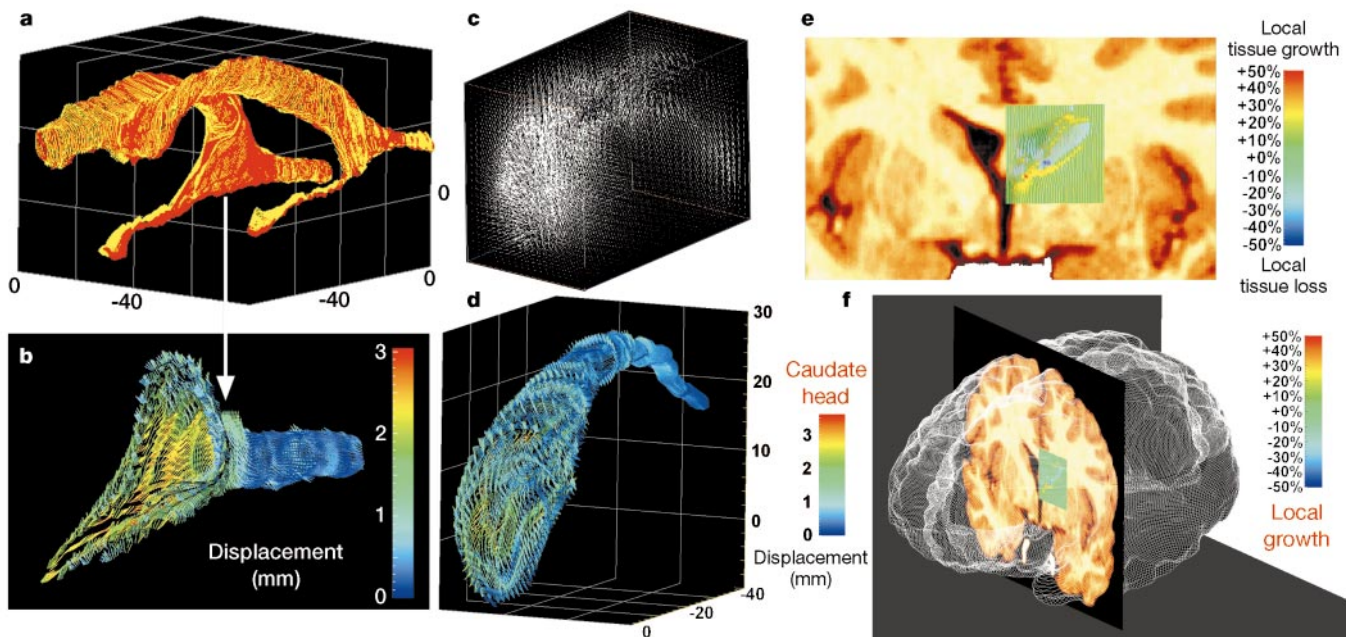


Figure 4 Detecting three-dimensional patterns of deep nuclear tissue loss. **a**, **b**, Tensor maps distinguish local growth or brain tissue loss from global displacements (**b**) of the adjacent ventricular anatomy, modelled here (**a**) at ages 7 years (red) and 11 years (yellow). **c**, **d**, Between ages 7 and 11 years, three-dimensional displacement vector maps show the deformation required to reconfigure earlier models of the caudate head

into their later shape. The caudate tail is stable (blue colours, **d**). **e**, **f**, Local growth (**e**) and anatomical displacement (**d**) of the caudate head are independently recovered, with 50% tissue loss detected locally (**e**, **f**), adjacent to a region of 20–30% growth throughout the internal capsule.

A subject scanned at ages 3 and 6 years exhibited a focus of peak growth rates (60–80% locally) throughout the anterior corpus callosum, in frontal circuits that help to sustain a vigilant mental state and regulate the organization and planning of new actions. The extremely rapid rates of local growth are consistent with metabolic studies using positron emission tomography⁵, which show an extraordinary doubling of the rates of glucose metabolism in the frontal cortex between ages 2 and 4 years, with frontal metabolic rates remaining at 199% of their adult values throughout the age range of 3–8 years. Between ages 3 and 6 years, when language function and associative thinking are not yet fully developed, growth rates at the isthmus were more quiescent (Fig. 2; 0–20% growth). Later growth foci in the isthmus, found consistently in all subjects aged 6–15 years, may reflect fine tuning of language functions known to occur late in childhood.

Regressive processes (tissue loss) were also detected at the same time as rapid growth. In the 7–11 and 9–13 year old subjects (Fig. 3), maps of lobar growth revealed pronounced (2–6 mm) temporo-parietal and pre-frontal enlargement. Somatosensory, motor and occipital brain regions were comparatively stable, with near-zero change in all brain regions at short time intervals (Fig. 3b). Up to 50% loss of tissue volume was detected at the caudate head (Fig. 4e and f). This tissue loss was highly localized, and contrasted with a 20–30% growth of the adjacent internal capsule (for which a separate surface model was made) and a 5–10% dilation of the superior ventricular horn (Fig. 4a). Gross volumetric measures confirmed an overall 60 mm³ tissue loss at the caudate head, although these global measures disguise the regional complexity of the change. This example helps illustrate how tensor maps distinguish local growth patterns (Fig. 4e) from bulk shifts, such as global displacements of the adjacent cerebral ventricles (Fig. 4a and b). Three-dimensional vector displacement maps (Fig. 4b and d) emphasize that both global and local displacements are required to match modelled anatomical elements across time. The three-dimensional deformation field, however, encodes the patterns of local anatomical dilation and contraction, and its values are unaffected by global displacements. Maps of local three-dimensional growth are therefore not critically dependent on how well scans are initially aligned, and can define growth at arbitrary three-dimensional points in the local anatomy (Fig. 4e). Figure 4f indicates the anatomical context and regional complexity of these growth and regressive processes. The foci of tissue loss corroborate the hypothesis that pruning processes occur during this developmental stage², suggesting that these processes can be tracked in an individual child.

We detected striking, spatially complex patterns of growth and tissue loss in the developing human brain. A rostro-caudal wave of peak growth rates (Fig. 1) was identified in the corpus callosum. Fibre systems that mediate language function and associative thinking grew more rapidly than surrounding regions across time spans before and during puberty (6–13 years), with growth attenuated shortly afterwards (11–15 years). This temporal pattern coincides with the ending of a well-known critical period for learning language, consistently noted in studies of second-language acquisition, including sign language, and in isolated children not exposed to language during early development⁶. The ability to learn new languages declines rapidly after the age of 12 years, as does the ability to recover language function if linguistic areas in one brain hemisphere are surgically resected. Peak growth rates in linguistic callosal regions, as well as their attenuation around puberty, may reflect the conclusion of the critical period for learning language and for accelerating signal transduction in networks that support both associative reasoning and language function. We recently found that the same temporo-parietal fibre system, crossing at the callosal isthmus, degenerates fastest in early Alzheimer's disease⁷, when progressive neuronal loss and perfusion deficits begin to occur in temporo-parietal association cortices and their commissural pro-

jection systems. The sensitivity of the approach may therefore offer advantages in tracking fine-scale effects of therapeutic interventions in dementia and oncology, mapping the local complexities of disease processes using dynamic rather than static criteria. □

Methods

Magnetic resonance imaging and pre-processing

Three-dimensional ($256^2 \times 124 \times 0.97$ mm $\times$ 0.97 mm $\times$ 1.5 mm resolution) T₁-weighted fast SPGR (spoiled GRASS (gradient-recalled acquisition in the steady state)) MRI volumes were acquired from young normal subjects (mean age 8.6 ± 3.1 years) at intervals ranging from two weeks to four years. For each scan pair, a radio-frequency bias field correction algorithm⁸ was applied to both scans to eliminate intensity drifts caused by scanner field inhomogeneity. The initial scan was then rigidly registered to the target using automated image registration software⁹ and resampled using chirp-Z (in-plane) and linear (out-of-plane) interpolation. Registered scans were histogram-matched (that is, their intensity distributions were equalized) and a preliminary map of differences in MRI signal intensities between the two scans was constructed¹⁰. Tensor models of structural change were then used to calculate rates of tissue dilation, contraction and shearing, mapping local patterns of change in three dimensions.

Image analysis

A high-resolution surface model of the cortex was automatically extracted¹¹ from each scan pair, and three-dimensional digital anatomical models, based on parametric surface meshes^{12,13}, were generated to represent a comprehensive set of deep sulcal, callosal, caudate and ventricular surfaces at each time point¹⁴. Surface models based on manually digitized data were averaged across multiple trials ($N = 6$) to minimize error¹⁵. These model surfaces provided anatomic constraints for an elastic image registration algorithm^{12,14}. For each subject, this algorithm calculated a three-dimensional elastic deformation vector field, with $384^2 \times 256 \times 3 \approx 0.1$ billion degrees of freedom, reconfiguring the anatomy at the earlier time point into the shape of the anatomy of the later scan. Surface deformations were used to derive a volumetric deformation field from which local measures of three-dimensional tissue dilation or contraction were quantified. Landmark points, surfaces, and curved anatomic interfaces were matched up in the pair of three-dimensional image sets, and the biological validity of the resulting anatomical transformation was guaranteed by forcing a large system of anatomical surface boundaries to match exactly. These included multiple structural, functional, and tissue type boundaries in three dimensions, including the callosum, caudate, cortex and ventricles. The deformation field driving the earlier onto the later anatomy was extended to the full volume using a continuum-mechanical model based on the Cauchy–Navier operator of linear elasticity^{14,16–18}. The resulting system of 0.1 billion second-order elliptic partial differential equations was solved by successive over-relaxation methods, with multi-grid acceleration^{12,14}, on a standard radiologic workstation. Potential artefactual differences due to differences in how surfaces were parameterized in each scan were compensated for, using a field of Christoffel symbols to modify the surface differential operators during the anatomical transformation¹⁴.

Tensor map computation

From this transformation, local rates of tissue dilation, contraction and shearing were calculated. Deformation processes recovered by the image-matching algorithm were analysed mathematically with vector field operators¹⁴ to produce a variety of tensor maps. These maps reflect the magnitude and principal directions of tissue dilation or contraction, and the local rates, divergence and gradients of the growth processes detected in the dynamically changing brain.

Received 27 August 1999; accepted 21 January 2000.

1. Fox, N. C., Freeborough, P. A. & Rossor, M. N. Visualisation and quantification of rates of atrophy in Alzheimer's disease. *Lancet* **348**, 94–97 (1996).
2. Giedd, J. N. *et al.* Quantitative magnetic resonance imaging of human brain development: ages 4–18. *Cereb. Cortex* **6**, 551–560 (1996).
3. Yakovlev, P. I. & Lecours, A. R. in *Regional Development of the Brain in Early Life* (ed. Minkowski, A.) 3–70 (Davis, Philadelphia, 1967).
4. Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L. & Toga, A. W. In vivo evidence for post-adolescent brain maturation frontal and striatal regions. *Nature Neurosci.* **2**, 859–861 (1999).
5. Chugani, H. T., Phelps, M. E. & Mazziotta, J. C. Positron emission tomography study of human brain functional development. *Ann. Neurol.* **22**, 487–497 (1987).
6. Grimshaw, G. M., Adelstein, A., Bryden, M. P. & MacKinnon, G. E. First-language acquisition in adolescence: evidence for a critical period for verbal language development. *Brain Lang.* **63**, 237–255 (1998).
7. Thompson, P. M. *et al.* Cortical variability and asymmetry in normal aging and Alzheimer's disease. *Cereb. Cortex* **8**, 492–509 (1998).
8. Zijdenbos, A. P. & Dawant, B. M. Brain segmentation and white matter lesion detection in MR images. *Crit. Rev. Biomed. Eng.* **22**, 401–465 (1994).
9. Woods, R. P., Cherry, S. R. & Mazziotta, J. C. Rapid automated algorithm for aligning and reslicing PET images. *J. Comp. Assist. Tomogr.* **16**, 620–633 (1992).
10. Freeborough, P. A., Woods, R. P. & Fox, N. C. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *J. Comp. Assist. Tomogr.* **20**, 1012–1022 (1996).
11. MacDonald, D., Avis, D. & Evans, A. C. in *Proc. SPIE Conf. Visualization in Biomedical Computing* (ed. Robb, R. A.) 2359, 160–169 (1994).

12. Thompson, P. M. & Toga, A. W. A surface-based technique for warping 3-dimensional images of the brain. *IEEE Trans. Med. Imag.* **15**, 471–489 (1996).
13. Thompson, P. M. & Toga, A. W. Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. *Med. Image Anal.* **1**, 271–294 (1997).
14. Thompson, P. M. & Toga, A. W. in *Brain Warping* (ed. Toga, A. W.) 311–336 (Academic, San Diego, 1998).
15. Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A. & Toga, A. W. 3D statistical analysis of sulcal variability in the human brain. *J. Neurosci.* **16**, 4261–4274 (1996).
16. Thompson, P. M. *et al.* Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *J. Comp. Assist. Tomogr.* **21**, 567–581 (1997).
17. Davatzikos, C. Spatial normalization of 3D brain images using deformable models. *J. Comp. Assist. Tomogr.* **20**, 656–665 (1996).
18. Miller, M. I. & Grenander, U. Computational anatomy: an emerging discipline. *Q. Appl. Math.* **56**, 617–694 (1998).

Correspondence and requests for materials should be addressed to P.M.T. (e-mail: thompson@loni.ucla.edu).

Acknowledgements

We thank E. Sowell, M. Mega and J. Mazzotta for their advice and support. P.M.T. was supported by the Howard Hughes Medical Institute, the US Information Agency, and the US–UK Fulbright Commission. Additional research support was provided by a Human Brain Project grant to the International Consortium for Brain Mapping, funded jointly by NIMH and NIDA, by National Institutes of Health intramural funding (J.N.G.), and by the National Library of Medicine, National Science Foundation, and the NCRR.

A clonogenic common myeloid progenitor that gives rise to all myeloid lineages

Koichi Akashi[†], David Traver^{*}, Toshihiro Miyamoto & Irving L. Weissman

Departments of Pathology and Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA

^{*} These authors contributed equally to this work

Haematopoietic stem cells give rise to progeny that progressively lose self-renewal capacity and become restricted to one lineage^{1,2}. The points at which haematopoietic stem cell-derived progenitors commit to each of the various lineages remain mostly unknown. We have identified a clonogenic common lymphoid progenitor that can differentiate into T, B and natural killer cells but not myeloid cells³. Here we report the prospective identification, purification and characterization, using cell-surface markers and flow cytometry, of a complementary clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Common myeloid progenitors give rise to either megakaryocyte/erythrocyte or granulocyte/macrophage progenitors. Purified progenitors were used to provide a first-pass expression profile of various haematopoiesis-related genes. We propose that the common lymphoid progenitor and common myeloid progenitor populations reflect the earliest branch points between the lymphoid and myeloid lineages, and that the commitment of common myeloid progenitors to either the megakaryocyte/erythrocyte or the granulocyte/macrophage lineages are mutually exclusive events.

The existence of clonal common lymphoid progenitors (CLPs)³ suggests that complementary progenitors common to all myeloid cells may also exist. Because the expression of the interleukin-7 receptor α -chain (IL-7R α) marks the CLPs and other downstream lymphoid progenitors^{3,4}, we searched the IL-7R α ⁺ fraction of

murine bone marrow for primitive myeloid progenitor populations.

In steady-state mouse bone marrow, myeloerythroid colony-forming unit (CFU) activity was found almost exclusively in the IL-7R α ⁺ Lin[−] c-Kit⁺ fraction (data not shown). Within this population, Sca-1⁺ cells are highly enriched for haematopoietic stem cells (HSCs)^{3,5–7}. To remove HSCs, Sca-1⁺ cells were excluded. The IL-7R α ⁺ Lin[−] c-Kit⁺ Sca-1[−] fraction was further divided into three subpopulations according to the expression profiles of the Fc γ receptor-II/III (Fc γ R), an important marker for myelomonocytic cells and a progenitor marker in fetal liver haematopoiesis⁸, and CD34, which marks a fraction of haematopoietic stem cells and progenitors⁶: the Fc γ R^{lo}CD34⁺, Fc γ R^{lo}CD34[−], and Fc γ R^{hi}CD34⁺ populations (Fig. 1a).

Each of the above populations were cleanly isolatable (Fig. 1b) and gave rise to distinct colony types in methylcellulose CFU assays (Figs 1c and 2). In the presence of steel factor (Slf), Flt-3 ligand (FL), IL-11, IL-3, granulocyte/macrophage-colony stimulating factor (GM-CSF), erythropoietin (Epo) and thrombopoietin (Tpo), ~80% of single multipotent HSCs randomly committed to myeloid lineages⁹, giving rise to various types of myeloid colonies including CFU-Mix¹⁰, burst-forming units-erythroid (BFU-E),

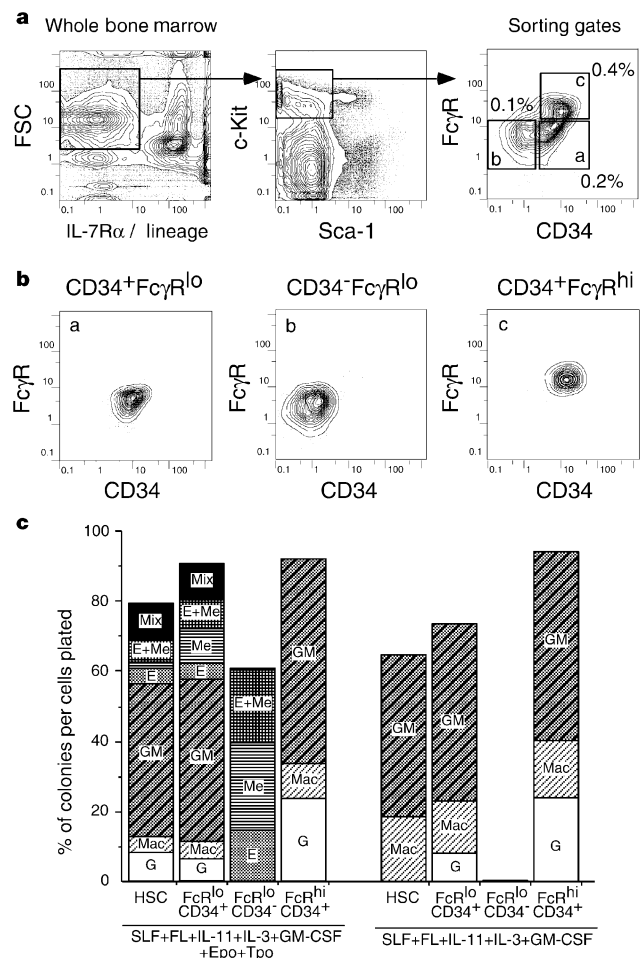


Figure 1 Identification of myeloid progenitors in mouse bone marrow. **a**, The IL-7R α ⁺ Lin[−] Sca-1[−] c-Kit⁺ fraction was subdivided into Fc γ R^{lo}CD34⁺, Fc γ R^{lo}CD34[−], and Fc γ R^{hi}CD34⁺ populations (a, b, c respectively as indicated in the right-hand panel). Percentages of each population relative to whole bone marrow are shown next to each sort gate. **b**, Re-analysis of the sorted Fc γ R^{lo}CD34⁺, Fc γ R^{lo}CD34[−] and Fc γ R^{hi}CD34⁺ populations. **c**, Clonogenic myeloid colony formation in methylcellulose. From each sorted progenitor population, 288 wells receiving a single cell each were scored. Fc γ R^{lo}CD34⁺ cells and HSCs formed various myeloid colonies including CFU-Mix, whereas the Fc γ R^{lo}CD34[−] and Fc γ R^{hi}CD34⁺ populations gave rise only to MegE and GM colonies, respectively (left). Megakaryocyte/erythroid colony formation from the Fc γ R^{lo} fractions was dependent upon Epo and/or Tpo (right).

[†] Present address: Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA.

CFU-megakaryocyte (CFU-Meg), CFU-megakaryocyte/erythroid (CFU-MegE), CFU-granulocyte/macrophage (CFU-GM), CFU-granulocyte (CFU-G) and CFU-macrophage (CFU-M), as we previously reported^{3,11}. More than 90% of sorted single $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^+$ cells gave rise to a colony distribution almost identical to that from HSCs. In contrast, more than 90% of $\text{Fc}\gamma\text{R}^{\text{hi}}\text{CD}34^+$ cells formed colonies composed only of macrophages and/or granulocytes such as CFU-M, CFU-G, or CFU-GM in response to any of the growth factor combinations, and are therefore termed granulocyte/macrophage lineage-restricted progenitors (GMPs). The $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^+$ cells gave rise exclusively to CFU-Meg, BFU-E or CFU-MegE colonies that contained only megakaryocytes and/or erythrocytes,

and are thus termed megakaryocyte/erythrocyte lineage-restricted progenitors (MEPs).

To verify the lineage restriction observed *in vitro*, we transplanted each of the three populations into lethally irradiated congenic recipient mice. Each of the progenitor populations rapidly differentiated into mature cells *in vivo*, and the cell fate outcomes strictly corresponded with those of *in vitro* colony assays. Six days after injection of 5,000 $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^+$ cells, both donor-derived $\text{Gr-1}^+/\text{Mac-1}^+$ myelomonocytic cells and TER119^+ erythroid cells were detectable in the spleen and bone marrow (data not shown). In contrast, 5,000 $\text{Fc}\gamma\text{R}^{\text{hi}}\text{CD}34^+$ GMPs transiently gave rise to only $\text{Gr-1}^+/\text{Mac-1}^+$ cells, whereas the $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^+$ MEPs reconstituted

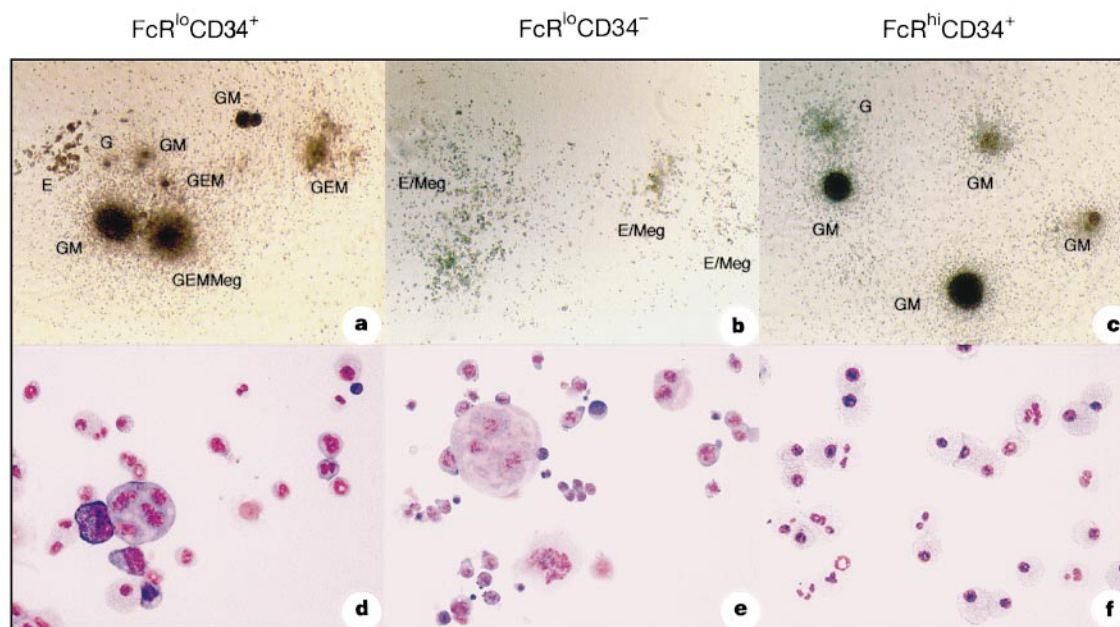


Figure 2 Morphology of day-7 colonies derived from sorted myeloid progenitors. Two-hundred cells from each population were cultured in methylcellulose containing Slf, IL-3, IL-11, GM-CSF, Epo, and Tpo in 35-mm dishes. The upper panels (a–c) show the

appearance and distribution of colonies, and the bottom panels (d–f) show the cellular morphology from five pooled colonies collected from each culture (Giemsa staining, original magnification $\times 1,000$).

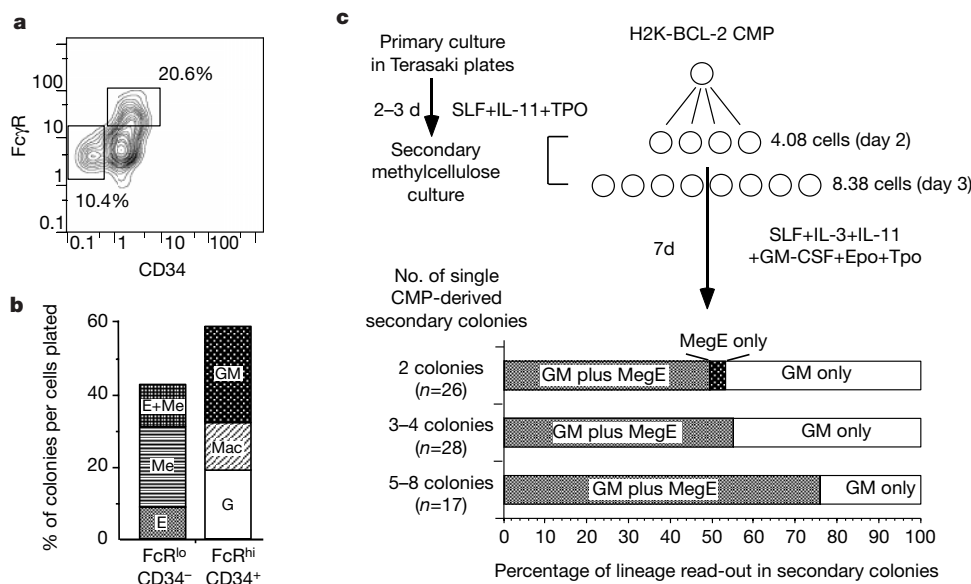


Figure 3 Lineage relationships among the myeloid progenitor subsets. **a**, $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^+$ CMPs gave rise to $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^+$ MEPs and $\text{Fc}\gamma\text{R}^{\text{hi}}\text{CD}34^+$ GMPs after 60-h cultures on S17 stromal layers in the presence of Slf. **b**, $\text{Fc}\gamma\text{R}^{\text{lo}}\text{CD}34^-$ MEPs and $\text{Fc}\gamma\text{R}^{\text{hi}}\text{CD}34^+$ GMPs purified from primary cultures generated colonies of megakaryocyte/erythroid (MegE) and granulocyte/macrophage (GM) lineages, respectively, in

methylcellulose containing Slf, IL-3, IL-11, GM-CSF, Epo and Tpo. **c**, Clonal analysis of CMP differentiation potential. Single CMPs were sorted into Terasaki wells containing liquid cultures supplemented with Slf, Tpo, and IL-11 and were expanded for 2–3 days. The number of cells in each well were then visualized and transferred to methylcellulose cultures (see Methods).

only TER119⁺ cells (data not shown). Myeloid progeny from each of these populations had disappeared by four weeks after injection, indicating that these progenitors have limited self-renewal capacity. These results are consistent with our previous finding that HSCs within whole bone marrow transplants are responsible for most white blood cells, platelets and red blood cells produced two weeks after transplant¹². We did not detect B-cell or T-cell progeny after either intravenous or intrathymic injections of $\approx 10,000$ MEPs or GMPs (data not shown). The Fc γ R^{lo}CD34⁺ population did not generate T cells. In a B-cell progenitor assay³, 1 in 2,780 cells differentiated into B cells (data not shown). Conversely, B-cell colony formation from CLPs in this assay was 1 in 8 cells (data not shown). Thus, the vast majority of cells within the Fc γ R^{lo}CD34⁺ population possess myeloid-restricted differentiation potential.

To test the lineage relationships among the three myeloid progenitor populations, we cultured 1,000 cells from each population on S17 stromal layers for 60 h in the presence of Slf and analysed their phenotypic changes by flow cytometry. Fc γ R^{lo}CD34⁺ cells gave rise to Fc γ R^{lo}CD34⁺ MEPs and Fc γ R^{hi}CD34⁺ GMPs (Fig. 3a). When Fc γ R^{lo}CD34⁺ cells and Fc γ R^{hi}CD34⁺ cells derived from cultured Fc γ R^{lo}CD34⁺ cells were re-sorted into methylcellulose, they formed colonies of megakaryocyte/erythrocyte (MegE) and granulocyte/macrophage (GM) lineages, respectively (Fig. 3b). In contrast, neither Fc γ R^{lo}CD34⁺ MEPs nor Fc γ R^{hi}CD34⁺ GMPs gave rise to the other two progenitor subtypes; progeny from these populations rapidly downregulated c-Kit expression and differentiated into mature cell types (data not shown). Thus, the Fc γ R^{lo}CD34⁺ cell is upstream of both Fc γ R^{lo}CD34⁺ MEP and Fc γ R^{hi}CD34⁺ GMP cells. On the basis of these data, we term Fc γ R^{lo}CD34⁺ cells common myeloid progenitors (CMPs).

We then sought to determine the lineage relationships among these three populations at the single-cell level. To increase the plating efficiency in this experiment, we purified CMPs from transgenic mice which have enforced expression of the anti-apoptotic *bcl-2* gene in all haematopoietic cells via the H-2^k Class I MHC promoter¹³. The *in vitro* myeloid CFU activity of each subset in

H2K-BCL-2⁺ bone marrow was identical to that in normal C57Bl bone marrow (data not shown). Single H2K-BCL-2 CMPs were sorted into Terasaki wells and cultured for 2–3 days in media containing the early acting cytokines Slf, Tpo, and IL-11, which stimulate proliferation with minimal differentiation. More than 90% of single H2K-BCL-2 CMPs divided once per day during the primary cultures and gave rise to ~ 4 and ~ 8 daughter cells by day 2 and day 3, respectively. From day 2 to day 3, between 4 and 16 daughter cells derived from single CMPs were transferred into secondary methylcellulose cultures containing Slf, FL, IL-11, IL-3, GM-CSF, Epo and Tpo (Fig. 3c). When the daughter cells derived from single CMPs gave rise to 2–4 secondary colonies, $\sim 55\%$ of these colonies consisted of both a GM-type colony such as CFU-GM, CFU-G or CFU-M, and a MegE-type colony such as BFU-E, CFU-E or CFU-MegE. When 5–8 secondary colonies were formed, 76% of these wells consisted of both GM- and MegE-type colonies. Upon averaging all secondary colonies, over 62% of single CMPs showed bipotential GM and MegE lineage outcomes in this clonal assay. CFU-Mix colonies were found in only 2 out of 264 secondary colonies ($< 1\%$). Because CMPs could give rise to $\sim 12\%$ CFU-Mix in conventional one-step methylcellulose cultures (Fig. 1c), most CMPs appear to lose multipotent myeloid differentiation activity during the primary suspension culture. This finding, in combination with the frequent coexistence of GM and MegE-related colonies in secondary cultures, strongly suggests that single CMPs do not appreciably self-renew but rapidly lose CMP activity and differentiate into GMPs and MEPs during the 2–3 cell divisions in primary culture.

The prospective isolation of HSCs, oligopotent progenitors and lineage-committed progenitors allows, for the first time to our knowledge, a definitive sampling of the transcriptional profiles of cells at distinct stages of differentiation and may allow the determination of whether selective expression (gain or loss) of specific genes is a cause or consequence of their commitment. Our study shows that the expression patterns of specific transcription factors within myeloid and lymphoid progenitors are generally consistent with

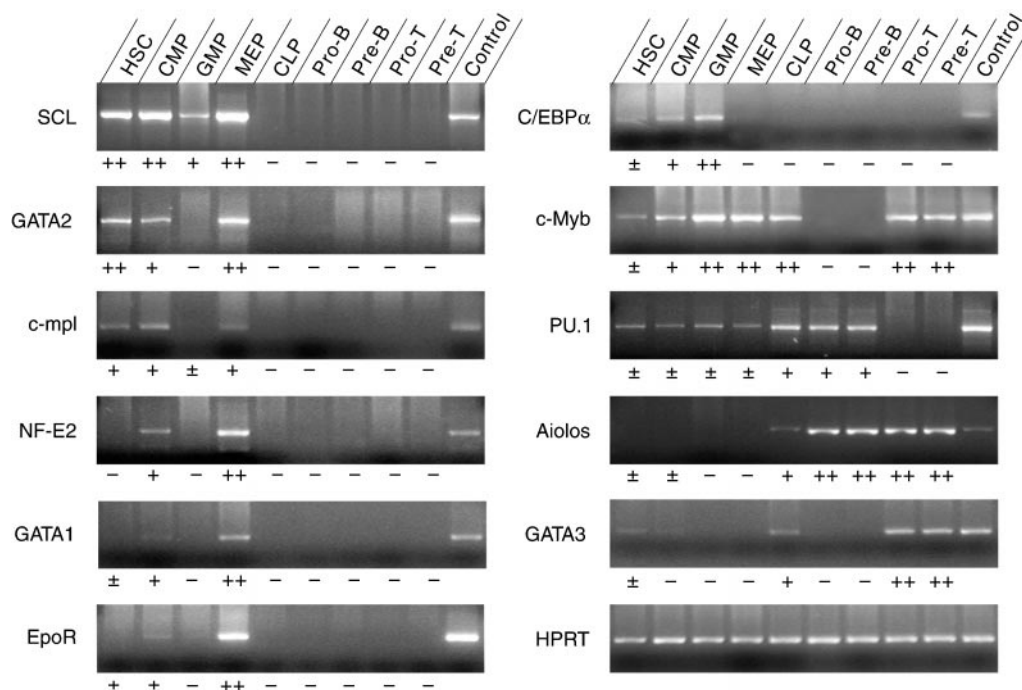


Figure 4 Differential expression of transcription factors in various stages of myeloid and lymphoid progenitors. Control cDNA was derived from 5×10^5 thymocytes for Aiolos, or 5×10^5 total bone marrow cells for remaining genes. The symbols under each lane depict relative amounts of messenger RNA in each population compared with control cDNA

(2×10^5 cells, bands not shown) by the ratio of Pixel Density Units of target cDNA to Pixel Density Units of control cDNA: less than 0.1 (-); 0.1–0.5 (±); 0.5–1.5 (+); more than 1.5 (++)

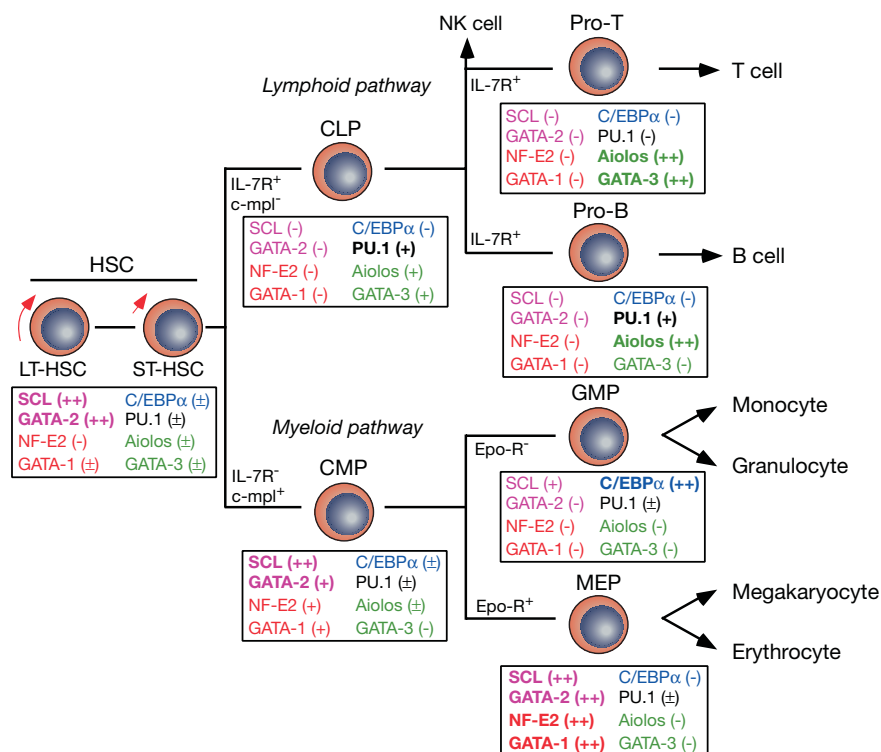


Figure 5 Proposed model of major haematopoietic maturation pathways from HSCs. Long-term HSCs (LT-HSCs) give rise to short-term HSCs (ST-HSCs). We propose that ST-HSCs give rise to at least CLPs, which can form all cells of the lymphoid lineages³, and CMPs, which can differentiate into either GMPs or MEPs which form the cells of the granulocyte/macrophage or megakaryocyte/erythroid lineages, respectively. These three

those expected from knockout studies (Fig. 4). For example, genes whose targeted disruptions lead primarily to defects in the megakaryocyte and/or erythroid lineages such as *GATA-1* and *NF-E2* are found to be most highly expressed in the MEP population. *GATA-1* is necessary in the post-commitment stages of erythroid and megakaryocyte development^{14–16}, and *NF-E2* is necessary for the normal maturation of megakaryocytes¹⁷. *C/EBPα* is required for granulocyte maturation¹⁸ and is likewise most highly expressed in the GMP population. Interestingly, *C/EBPα*, *NF-E2* and *GATA-1* are co-expressed at relatively low levels in CMPs. In parallel, similarly low-level expression of lymphoid-related transcription factors such as *Pax-5* (ref. 19) (data not shown), *Aiolos*²⁰ and *GATA-3* (ref. 21) is found in CLPs but not in any of the myeloid progenitor populations. These data support intrinsically restricted differentiation potentials of each respective progenitor population through differential gene expression programs¹⁸. Furthermore, low-level expression of transcription factors in progenitors common to the myeloid and lymphoid lineages may reflect priming stages at which oligo-lineage commitment remains flexible^{22,23}. Figure 5 shows our proposed model of the major haematolymphoid maturation pathways from HSCs.

Our results, together with previous work on HSC subsets^{6,7} and CLPs³, provide a means to isolate cells at several stages of haematopoietic differentiation, including at least three critical decision points: (1) the loss of self-renewal potential as long-term HSCs transit to and through short-term HSCs; (2) the decision of HSCs to choose the lymphoid lineage (by generating CLPs) or the myeloid lineage (by generating CMPs); and (3) the decision of CMPs to follow the granulocyte/macrophage lineage (by generating GMPs) or the megakaryocyte/erythrocyte lineage (by generating MEPs). The ability to isolate each population prospectively should allow an enhanced ability to identify candidate genes involved in lineage commitment, and will allow the transduction of these genes into

myeloid progenitor subsets should comprise the vast majority of myeloid progenitor activity in steady-state bone marrow, because the Sca-1⁺c-Kit⁺ fraction, which is composed of GMPs, MEPs and CMPs, and the Sca-1⁺c-Kit⁺ HSC fraction were estimated to contain ~98% of myeloid colony-forming activity within the Lin[−]IL-7Rα[−] fraction (data not shown).

isolated progenitors to resolve their roles. These lineage-committed precursors should also help clarify the currently controversial origins of mast cells, basophils, eosinophils and 'lymphoid' versus 'myeloid' dendritic cells. Finally, using mouse and human models of various types of leukaemias, we hope to test whether transformation events leading to clonogenic transplantable leukaemias occur within these defined stem or progenitor cell populations, or whether the leukaemias themselves are more differentiated cells that have acquired the proliferative and self-renewal potentials of stem cells. □

Methods

Mouse strains

The congenic strains of mice, C57BL/Ka-Thy1.1 (Ly5.1) and C57BL/Ka-Thy1.1-Ly5.2, were used as described³. H2K-BCL-2 mice have been described¹³. All animals were maintained in Stanford University's Research Animal Facility in accordance with Stanford guidelines.

Cell staining and sorting

For myeloid progenitor experiments, bone marrow cells were stained with biotinylated antibodies specific for the following lineage markers: CD3 (KT31.1), CD4 (GK1.5), CD8 (53-6.7), B220 (6B2), Gr-1 (8C5), TER119, CD19 (1D3), IgM (R6-60.2, Pharmingen) and IL-7Rα chain (A7R34). Lin⁺ cells were partially removed with sheep anti-rat IgG-conjugated magnetic beads (Dynabeads M-450, Dynal A.S.), and the remaining cells were stained with avidin-Cy5-PE (Tricolor, Caltag). Cells were stained with PE-conjugated anti-FcγRII/III (2.4G2), FITC-conjugated CD34 (RAM34) (Pharmingen), Texas red-conjugated anti-Sca-1 (E13-161-7) and APC-conjugated anti-c-Kit (2B8) monoclonal antibodies. Sorting methods for HSCs, CLPs and proT and proB cells have been described³. All cell populations were sorted or analysed using a highly modified triple laser (488-nm argon laser, 599-nm dye laser and ultraviolet laser) FACS Vantage (Becton Dickinson Immunocytometry Systems).

In vitro and in vivo assays to determine differentiation potential of progenitors

To support the formation of myeloid colonies, progenitors were cultured in an alpha-Modified Eagle Medium (αMEM)-based methylcellulose media (Methocult M3100; StemCell Technologies, Vancouver) that was supplemented as described¹¹. CFU-Mix such as CFU-GEMMeg, CFU-GEM and CFU-GEMeg were scored by Giemsa staining of

individual colonies plucked using fine-drawn Pasteur pipettes. To evaluate the lineage relationships among CMPs, MEPs and GMPs, 1,000 cells of each population were cultured for 60 h on irradiated (3,000 rad) S17 stromal layers in 24-well plates with RPMI 1640 medium containing 10% FBS (Gemini Bioproducts) and SLf (10 ng ml⁻¹). A two-step culture assay was also carried out to show multilineage differentiation capacity of single CMPs. Single CMPs from H2k-BCL-2 mice were deposited into Terasaki wells containing 15 µl of IMDM supplemented with 10% FCS, SLf (20 ng ml⁻¹), Tpo (10 ng ml⁻¹), and IL-11 (20 ng ml⁻¹) using an ACDU cloning system. After 2–3 days in culture, 4–16 single cell-derived daughter cells were transferred into methylcellulose containing SLf, IL-3, IL-11, GM-CSF, Tpo and Epo, and the secondary colonies were enumerated after 7–10 days. The *in vitro* B-cell differentiation potential of CMPs and CLPs was done as described³ by using OP9 stromal layers. All cultures were incubated at 37 °C in a humidified chamber under 7% CO₂. For reconstitution assays, purified progenitors were injected into the retro-orbital sinuses of lethally irradiated (920 rad delivered in a split dose) congenic mice (differing only at the Ly5 allele) with or without 200 host-type HSCs.

Evaluation of transcription factor expression

Total RNA was purified from 1,000 double-sorted cells from each population, and was amplified by RT-PCR as previously reported²⁴. Primer sequences were as previously reported^{20,25–30}. Quantification of each message was carried out by comparing the expression levels of test samples to control complementary DNAs prepared from 2 × 10⁵ whole bone marrow cells or thymocytes, using the Integrated Image analysis system (Bio-Rad Laboratories). Polymerase chain reaction cycle number for each target gene was optimized to obtain linear correlations between pixel density units of test and control cDNAs. Polymerase chain reaction products were visualized by a Gel Doc 1000 Video Gel Documentation System and quantitated by Molecular Analyst Software (Bio-Rad).

Received 25 November 1999; accepted 19 January 2000.

1. Dexter, T. M. Introduction to the haemopoietic system. *Cancer Surv.* **9**, 1–5 (1990).
2. Metcalf, D. Stem cells, pre-progenitor cells and lineage-committed cells: are our dogmas correct? *Ann. NY Acad. Sci.* **872**, 289–303 (1999).
3. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
4. Akashi, K., Kondo, M., von Freeden-Jeffry, U., Murray, R. & Weissman, I. L. Bcl-2 rescues T lymphopoiesis in interleukin-7 receptor-deficient mice. *Cell* **89**, 1033–1041 (1997).
5. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
6. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34⁺ low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
7. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
8. Lacaud, G., Carlsson, L. & Keller, G. Identification of a fetal hematopoietic precursor with B cell, T cell, and macrophage potential. *Immunity* **9**, 827–838 (1998).
9. Ogawa, M. Differentiation and proliferation of hematopoietic stem cells. *Blood* **81**, 2844–2853 (1993).
10. Metcalf, D., Johnson, G. R. & Mandel, T. E. Colony formation in agar by multipotential hemopoietic cells. *J. Cell Physiol.* **98**, 401–420 (1979).
11. Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. *Nature Med.* **2**, 1011–6 (1996).
12. Uchida, N., Aguila, H. L., Fleming, W. H., Jerabek, L. & Weissman, I. L. Rapid and sustained hematopoietic recovery in lethally irradiated mice transplanted with purified Thy-1.1lo Lin[−]Sca-1⁺ hematopoietic stem cells. *Blood* **83**, 3758–3779 (1994).
13. Domen, J., Gandy, K. L. & Weissman, I. L. Systemic overexpression of BCL-2 in the hematopoietic system protects transgenic mice from the consequences of lethal irradiation. *Blood* **91**, 2272–2282 (1998).
14. Pevny, L. *et al.* Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260 (1991).
15. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl Acad. Sci. USA* **93**, 12355–12358 (1996).
16. Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J.* **16**, 3965–3973 (1997).
17. Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. Absence of blood formation in mice lacking the T-cell leukaemia oncogene tal-1/SCL. *Nature* **373**, 432–434 (1995).
18. Tenen, D. G., Hromas, R., Licht, J. D. & Zhang, D. E. Transcription factors, normal myeloid development, and leukemia. *Blood* **90**, 489–519 (1997).
19. Nutt, S. L., Urbanek, P., Rolink, A. & Busslinger, M. Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes Dev.* **11**, 476–491 (1997).
20. Morgan, B. *et al.* Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**, 2004–2013 (1997).
21. Ting, C. N., Olson, M. C., Barton, K. P. & Leiden, J. M. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* **384**, 474–478 (1996).
22. Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785 (1997).
23. Enver, T. & Greaves, M. Loops, lineage, and leukemia. *Cell* **94**, 9–12 (1998).
24. Miyamoto, T. *et al.* Persistence of multipotent progenitors expressing AML1/ETO transcripts in long-term remission patients with t(8;21) acute myelogenous leukemia. *Blood* **87**, 4789–4796 (1996).
25. Weiss, M. J., Keller, G. & Orkin, S. H. Novel insights into erythroid development revealed through *in vitro* differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **8**, 1184–1197 (1994).
26. Robb, L. *et al.* Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc. Natl Acad. Sci. USA* **92**, 7075–7079 (1995).
27. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2**, 786–800 (1988).

28. Ko, L. J. *et al.* Murine and human T-lymphocyte GATA-3 factors mediate transcription through a cis-regulatory element within the human T-cell receptor delta gene enhancer. *Mol. Cell. Biol.* **11**, 2778–2784 (1991).
29. Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M. V. Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol. Cell. Biol.* **13**, 473–486 (1993).
30. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P. & Orkin, S. H. Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362**, 722–728 (1993).

Acknowledgements

This work was supported by a USPHS US Public Health Service grant to I.L.W. and a 1997 Jose Carreras International Leukemia Foundation grant to K.A. D.T. is supported by a National Institute of Allergy and Infectious Diseases Training Grant. We thank S.-I. Nishikawa for anti-IL-7R antibody, J. Domen for H2K-BCL-2 mice and helpful discussions, D. Wright and A. Kiger for critical evaluation of the manuscript, L. Jerabek for excellent laboratory management and assistance with animal procedures, V. Braunstein for antibody preparation, the Stanford FACS facility for flow cytometer maintenance, and L. Hidalgo and B. Lavarro for animal care.

Correspondence and requests for materials should be addressed to K.A. (e-mail: akashi@leland.stanford.edu).

Regulation of intracellular calcium by a signalling complex of IRAG, IP₃ receptor and cGMP kinase Iβ

Jens Schlossmann*, Aldo Ammendola*†, Keith Ashman*‡, Xiangang Zong*, Andrea Huber*, Gitte Neubauer‡, Ge-Xin Wang§, Hans-Dieter Allescher*||, Michael Korth§, Matthias Wilm‡, Franz Hofmann* & Peter Ruth*

* Institut für Pharmakologie und Toxikologie der Technischen Universität München, Biedersteiner Straße 29, 80802 München, Germany

‡ Protein and Peptide Group, EMBL, Meyerhofstraße 1, 69117 Heidelberg, Germany

§ Abteilung Pharmakologie für Pharmazeuten, Universitäts-Krankenhaus Eppendorf, Martinistraße 52, 20246 Hamburg, Germany

† These authors contributed equally to this work

Calcium release from the endoplasmic reticulum controls a number of cellular processes, including proliferation and contraction of smooth muscle and other cells^{1,2}. Calcium release from inositol 1,4,5-trisphosphate (IP₃)-sensitive stores is negatively regulated by binding of calmodulin to the IP₃ receptor (IP₃R)^{3,4} and the NO/cGMP/cGMP kinase I (cGKI) signalling pathway^{5,6}. Activation of cGKI decreases IP₃-stimulated elevations in intracellular calcium⁷, induces smooth muscle relaxation⁸ and contributes to the antiproliferative⁹ and pro-apoptotic effects of NO/cGMP¹⁰. Here we show that, in microsomal smooth muscle membranes, cGKIβ phosphorylated the IP₃R and cGKIβ, and a protein of relative molecular mass 125,000 which we now identify as the IP₃R-associated cGMP kinase substrate (IRAG). These proteins were co-immunoprecipitated by antibodies directed against cGKI, IP₃R or IRAG. IRAG was found in many tissues including aorta, trachea and uterus, and was localized perinuclearly after heterologous expression in COS-7 cells. Bradykinin-stimulated calcium release was not affected by the expression of either IRAG or cGKIβ, which we tested in the absence and presence of cGMP. However, calcium release was inhibited after co-expression of IRAG and cGKIβ in the presence of cGMP. These results identify IRAG as an essential NO/cGKI-dependent regulator of IP₃-induced calcium release.

|| Present address: II. Medizinische Klinik und Poliklinik der TU München, Ismaninger Str. 22, D-81675 München, Germany.

Proteins with relative molecular masses of 240,000, 125,000 and 80,000 (M_r 240K, 125K and 80K), termed G_0 , G_1 and G_2 , have been identified as substrates of cGKI in vascular smooth muscle membranes¹¹; G_0 has been identified as IP₃R¹². The IP₃R is phosphorylated by cGKI¹³, but no significant function¹⁴ or Ca²⁺-release sensitizing¹⁵ or inhibiting¹⁶ effects have been associated with this phosphorylation. In contrast to G_0 , the identity of G_1 and G_2 has remained speculative^{17,18}.

We found that microsomal membranes from bovine tracheal smooth muscle contained G_0 , G_1 and G_2 , together with the substrates G_3 and G_4 which were phosphorylated in the presence of cGMP (Fig. 1a). Antibodies specific for the IP₃R and cGKI β labelled G_0 and G_2 in immunoblots, respectively (Fig. 1b). The membrane proteins were solubilized and the soluble proteins were immunoprecipitated by these antibodies and also by an antibody that reacted with G_1 . All three antibodies co-precipitated the IP₃R (G_0), IRAG (G_1) and cGKI β (G_2) to a similar extent (Fig. 1c), suggesting that these proteins formed a multimeric complex. As reported previously¹⁸, the myosin-binding subunit, which only interacts with cGKI α ¹⁹, was not detected in this complex.

The presence of cGKI β allowed the purification of the solubilized complex by cGMP-agarose (Fig. 2a). The purified and phosphorylated complex proteins were separated on preparative SDS polyacrylamide gel electrophoresis (PAGE). Substantial amounts of G_1 (Fig. 2a), G_0 and G_2 were purified for protein sequencing (Fig. 2a). The matrix-assisted laser desorption ionization (MALDI) spectrum of G_0 and G_2 identified G_0 as type 1 IP₃R and G_2 as cGKI β , whereas the MALDI spectrum of G_1 was not matched by any known protein. Amino-acid sequencing yielded several peptides including four phosphopeptides (Fig. 2b). An expressed sequence tag (EST)-based nucleotide sequence comprising peptides 2–7 (Fig. 2b) was used to screen an oligo dT-primed library from tracheal muscle, yielding several poly(A)-tailed clones. Additional screening of a specifically primed library completed the complementary DNA sequence. The hitherto unidentified cGKI substrate G_1 , now termed IRAG (Fig. 2c), is a new protein. Sequence alignment showed that the sequence from residue 342 to residue 911 is 42% homologous to Jaw1, a lymphoid-restricted membrane protein of the endoplasmic reticulum²⁰. Analysis of several amino-terminal sequences yielded four different transcripts coding for IRAGa, IRAGb and the N-terminal deletion forms Δ IRAGa and Δ IRAGb (Fig. 2c). Using a specific N-terminal primed library, the number of identified clones were 3, 9, 4 and 9 for IRAGa, IRAGb, Δ IRAGa and Δ IRAGb, respectively. Δ IRAGa and Δ IRAGb coded for identical

proteins, but differed in their 5' untranslated region (UTR) (Fig. 2c). Both clones lacked the exon containing the start ATG of IRAGa and IRAGb. Translation of these clones could be initiated at the methionine (CTG) codon. A similar initiation site has been suggested for splice variants of the neuronal NO synthase²¹. Hydrophathy analysis of IRAG yielded a single carboxy-terminal membrane-spanning segment at the N terminus. The central part of the protein exhibited a coiled-coil structure allowing protein–protein interactions. The identified *in vitro* phosphorylation sites for cGKI β

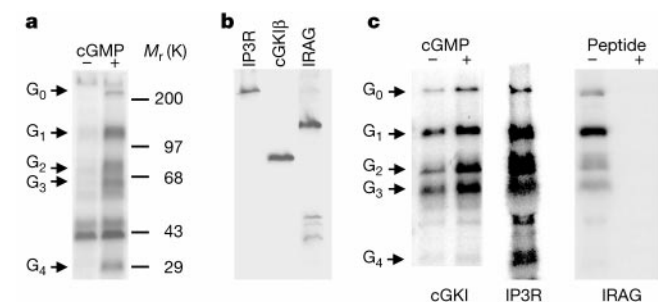


Figure 1 Macromolecular complex of IP₃R, IRAG and cGKI β . **a**, Autoradiogram of microsomal membranes from tracheal smooth muscle phosphorylated in the absence or presence of cGMP. The substrates G_0 – G_4 of cGKI β are indicated. **b**, Immunoblots using IP₃R, cGKI β and IRAG specific antibodies. The lower bands are nonspecific staining observed also with preabsorbed antibody. **c**, Immunoprecipitation of solubilized proteins by antibodies specific for cGKI (left), IP₃R (middle) and IRAG (right). The antibody for IRAG was preincubated with (+ peptide) or without (– peptide) the peptide used for immunization. Autoradiograms of immunoprecipitated and thereafter phosphorylated proteins are shown.

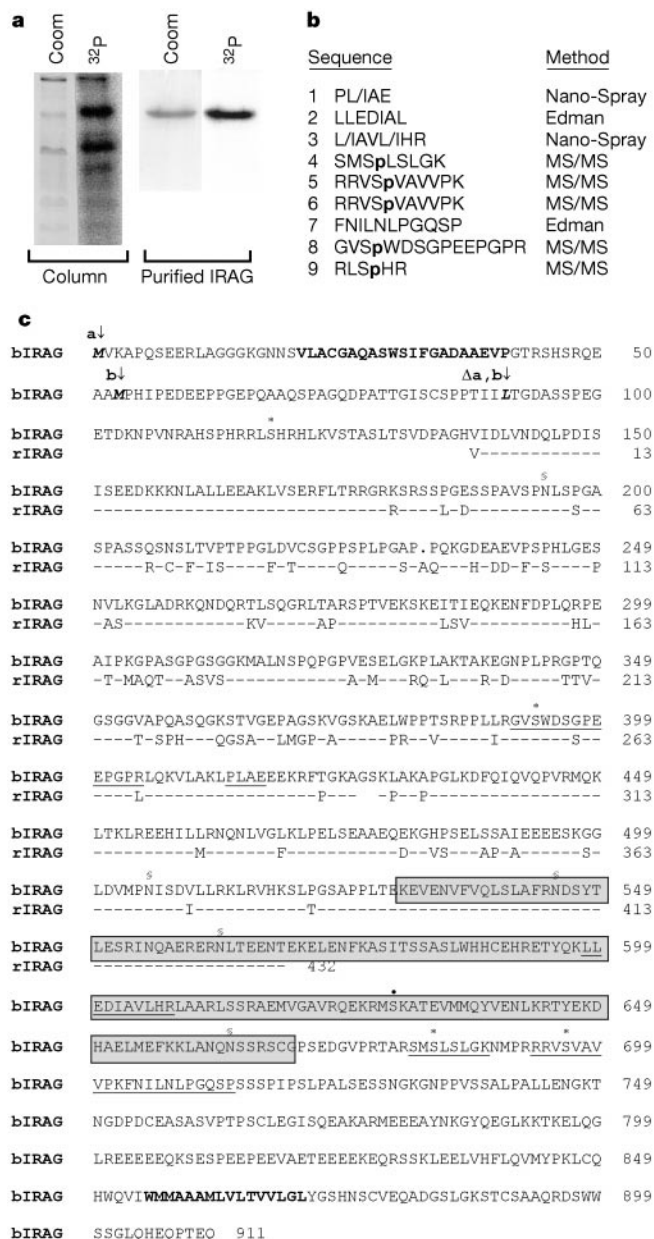


Figure 2 Identification of IRAG. **a**, Microsomal membranes were solubilized and purified on cGMP agarose. The phosphorylated proteins were analysed by SDS–PAGE. Purified complex (left); purified IRAG (right). Coom, Coomassie staining; ³²P, autoradiogram. **b**, Tryptic peptides of IRAG (underlined in **c**). Sp, phospho-serine. **c**, Protein sequence of IRAG from bovine tracheal muscle (bIRAG). The translation start sites of IRAGa, IRAGb and Δ IRAGa,b at amino-acid position 1, 53 and 91, respectively, are indicated in bold and italic. Putative transmembrane helices are indicated in bold. The coiled-coil structure is boxed. Potential N-glycosylation (S), identified (*) and putative (●) phosphorylation sites for cGMP-kinase are indicated. rIRAG, a partial sequence of rat IRAG is aligned to bIRAG. Identical amino acids are given as a dash.

are N-terminal and C-terminal of the coiled-coil domain, implying that regulatory function resides in this part of the molecule which is exposed to the cytosol.

In northern blots, a single messenger RNA species of 6.5 kilobases (kb) was detected in tracheal smooth muscle and other tissues (Fig. 3a, b). This size is consistent with the length of coding (2.8 kb), 5'-UT (0.3 kb) and 3'-UT (3.4 kb) region obtained by cloning. Western blots using an antibody raised against the common sequence of IRAG revealed high expression in trachea, aorta (Figs 1b and 3c) and uterus. Expression of IRAGa and IRAGb in COS-7 cells that did not express IRAG endogenously yielded recombinant proteins with identical (IRAGb; M_r 125K) or similar (IRAGa; M_r 130K) molecular mass to native IRAG (Fig. 3c). Like the native protein (Fig. 1c), recombinant IRAG formed a complex with cGKI β when both cDNAs were transfected into COS-7 cells (Fig. 3c). Using polymerase chain reaction with reverse transcription (RT-PCR) analysis, we confirmed that COS-7 cells contain low amounts of type 1 IP $_3$ R²². Owing to its low density, the IP $_3$ R was not detected in the complex. To confirm that IRAG and cGKI β interacted with each other, a two-hybrid screen was carried out using cGKI β as bait and an intestinal smooth muscle cDNA library. This yielded several positive clones which were identified as IRAG (Fig. 2c). The interacting fragment of IRAG contains part of the coiled-coil domain, indicating that cGKI β may interact with IRAG through this domain. The intracellular localization of IRAG was studied by

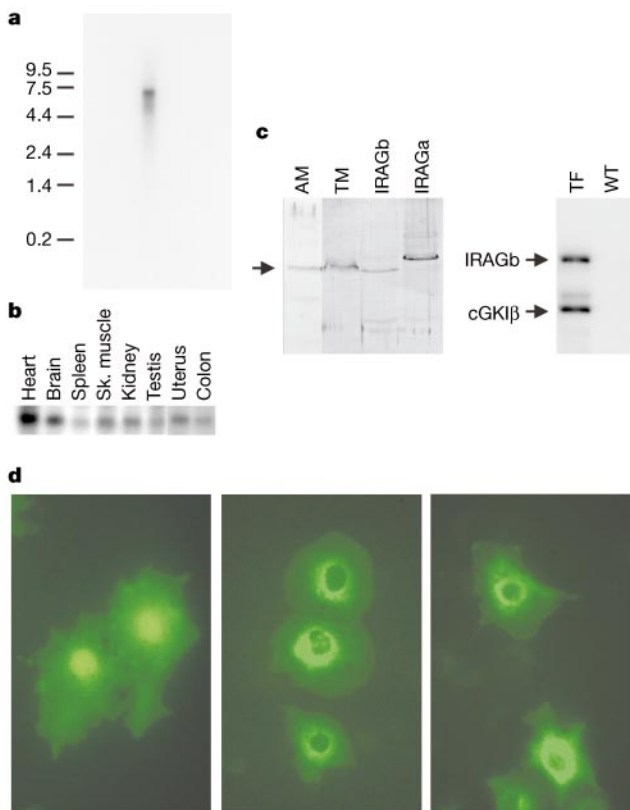


Figure 3 Expression and localization of IRAG. **a**, Northern blot performed with mRNA from bovine tracheal muscle and a specific probe derived from the common region of IRAG. **b**, Northern blot of several human tissues (Clontech) probed with an amplified fragment of human IRAG (EST m78793). **c**, Left, immunoblot of bovine aorta (AM) and trachea (TM) microsomal membranes and COS-7 cells transfected with bIRAGa or bIRAGb using the IRAG specific antibody. Right, COS-7 cells co-expressing bIRAGb-GFP fusion protein and cGKI β were solubilized. Proteins bound to cGMP agarose were phosphorylated, eluted and analysed by autoradiography (TF). Mock transfected cells (WT) were used as a control. **d**, Fluorescence micrographs of cells expressing GFP (left), bIRAGb-GFP (middle) and bIRAGa-GFP (right) (original magnification $\times$ 400).

fusing IRAGa and IRAGb to green fluorescent protein (GFP) at their N termini. As expected for a protein interacting with the IP $_3$ R, IRAGa and IRAGb fluorescence was localized perinuclearly, whereas free GFP accumulated in the nucleus (Fig. 3d) supporting the idea that IRAG, IP $_3$ R and cGKI β formed a complex in the endoplasmic reticulum. This perinuclear localization is of interest because the elementary events in HeLa cells that result from Ca²⁺ release from IP $_3$ Rs are located mostly near the nucleus²³.

To investigate the potential cellular function of IRAG, the IRAG-GFP constructs were expressed alone or together with cGKI β in COS-7 cells. The release of Ca²⁺ from the endoplasmic reticulum was stimulated by bradykinin, and the calcium transients were monitored by Fura-2 fluorescence. COS-7 cells were selected because they express type 1 IP $_3$ R, but not IRAG or cGKI β proteins. Transfection of the cells with the various constructs did not significantly affect basal and bradykinin-stimulated Ca²⁺ transients. To analyse the effect of cGMP-dependent phosphorylation of IRAG or IP $_3$ R on the Ca²⁺ transients, the cells were stimulated with bradykinin, pre-incubated in the absence and presence of the cGK-specific cGMP analogue 8-pCPT-cGMP, and then stimulated again with bradykinin (Fig. 4). For comparison, the area under the transient curve (AUC) was calculated. This value reflects the total amount of calcium released from the IP $_3$ -sensitive store. The second transients were not significantly different from the first transients in cells transfected with cGKI β , IRAGa or IRAGb alone, whether the cells were pre-incubated with cGMP or not (Fig. 4, Table 1). The Ca²⁺ transients were not significantly affected in cells expressing cGKI β activated with 8-pCPT-cGMP when the pipette concentration of bradykinin was lowered to 30 nM. In contrast, transients were suppressed significantly when IRAGa or IRAGb was co-expressed with cGKI β and the kinase was activated by pre-incubation with 8-pCPT-cGMP. cGKI β did not reduce Ca²⁺ transients in cells expressing IRAG in the absence of 8-pCPT-cGMP, indicating that activation of cGKI β was required.

A potential interpretation of these results was that activation of cGKI β interfered with the synthesis of IP $_3$, as has been shown for

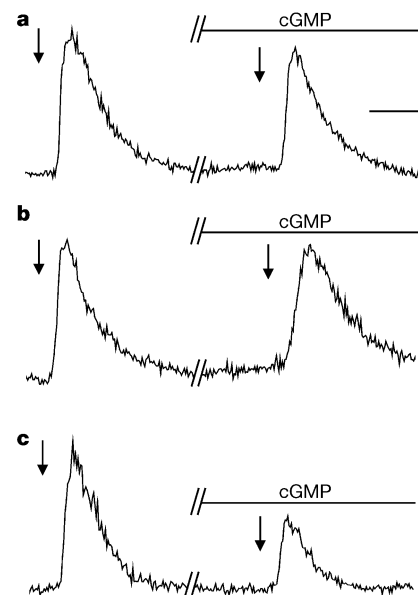


Figure 4 IRAG attenuates bradykinin-induced Ca²⁺ release. COS-7 cells were transfected with cGKI β (**a**), bIRAGa (**b**) or bIRAGa and cGKI β (**c**). Cells were stimulated with bradykinin (100 nM in the pipette) for 10 s (first transient). The cells were then superfused with buffer for 10 min (gap), the last 5 min with or without 8-pCPT-cGMP (100 μ M), followed by a second stimulation with bradykinin. The arrow marks the start of the bradykinin superfusion. The horizontal and vertical bars indicate 20 s and 200 nM Δ [Ca²⁺]_i, respectively.

cGKI α ⁷. However, this was not the case (Fig. 5). Perfusion of COS-7 cells with IP₃ released Ca²⁺ from intracellular stores in the absence of cGMP, in wild-type cells and cells transfected with IRAG α and cGKI β . Pre-incubation of the transfected but not the wild-type cells with 8-pCPT-cGMP almost completely suppressed the IP₃-induced Ca²⁺ release, supporting a direct interaction between cGKI β and IRAG and ruling out that phosphorylation of IRAG reduced the IP₃ synthesis. As cGKI β is the main cGKI isoform of smooth muscle²⁴, these results strengthen the idea that the activity of cGKI β is an important parameter to control smooth muscle tone. This does not preclude additional mechanisms, such as the phosphorylation of the myosin-binding subunit of myosin phosphatase by cGKI α ¹⁹, contributing to this regulation.

These results show that the NO/cGMP/cGKI pathway regulates IP₃-mediated Ca²⁺ release by the phosphorylation of the protein IRAG. We can not exclude the possibility that the decreased Ca²⁺ release requires cGMP-dependent phosphorylation of the IP₃R and IRAG; however, the results show that phosphorylation of the IP₃R is

not sufficient to affect Ca²⁺ release. IRAG may affect the dynamics of Ca²⁺ release either by modulating IP₃R function or by regulating systems that provide electroneutrality of Ca²⁺ release^{25,26}. In smooth muscle, phosphorylation of IRAG by cGKI is probably a major mechanism that reduces intracellular calcium and relaxes vascular tone. The importance of IRAG extends to the regulation of growth in other cells, as it has been reported that the murine AIDS-related virus integrates into the MRV1 gene, an IRAG homologue, leading to myeloid leukaemia²⁷. The level of intracellular calcium controls the life and death of a cell (see ref. 1 for review); it is therefore conceivable that IRAG is a molecule that integrates signals from different pathways by modulating the release of calcium from the endoplasmic reticulum. □

Methods

Preparation and phosphorylation of tracheal smooth muscle membranes

Bovine tracheal smooth muscle was prepared and homogenized in 20 mM MOPS pH 7.4, 8% sucrose. The homogenate was filtered through cheesecloth. The filtrate was centrifuged at 10,000g for 10 min at 4 °C. The supernatant was centrifuged at 130,000g for 60 min at 4 °C. The resulting pellet was resuspended in 20 mM MOPS pH 7.4, 160 mM NaCl. Membranes (30 μ g) were incubated in 50 mM Mes pH 6.9, 10 mM NaCl, 1 mM MgAc, 0.4 mM EGTA, 0.1 mM [γ -³²P]ATP (2,000 c.p.m. per pmol) in the presence or absence of 8-pCPT-cGMP (3 μ M) for 2 min at 30 °C. The reactions were stopped by addition of Laemmli buffer. Proteins were separated by SDS-PAGE and phosphorylation was analysed by autoradiography.#

Co-immunoprecipitation

Specific antibodies (7–20 μ l) directed against cGKI β , IP₃R type 1 (ABR Biochemicals) or bIRAG (residues 93–107) were pre-bound to protein A Sepharose beads. Microsomal membrane proteins (200 μ g) were solubilized in 4 mM MOPS pH 7.4, 32 mM NaCl, 1.6% lubrol-PX (20 min, 4 °C), centrifuged (11,000g, 10 min, 4 °C) and added to the beads. After incubation for 2 h at 4 °C, the beads were washed and incubated in phosphorylation buffer in the presence or absence of 8-pCPT-cGMP (2 min, 30 °C). Proteins were eluted with Laemmli buffer and analysed by SDS-PAGE and autoradiography.

Purification of cGKI β substrates

Microsomal membranes (400 mg), prepared from 600 g muscle of 60 bovine tracheas, were solubilized and incubated with 2.5 ml 8-AET-cGMP agarose beads (Biolog) for 90 min. The beads were washed and incubated with phosphorylation buffer (2 min, 30 °C) in the presence of 8-pCPT-cGMP (3 μ M). Proteins were eluted by addition of Laemmli buffer and analysed by SDS-PAGE and autoradiography. Phosphorylated protein corresponding to G₀, G₁ and G₂ substrate was excised from the gel and eluted with 25 mM Tris, 192 mM glycine, 0.1% SDS, using an electro-eluter (Bio-Rad). The yield of G₁ was 40 μ g. G₁ was not identical with the 130K subunit of myosin light chain phosphatase I as shown by immunoblots using the specific antibodies F38 and M110, provided by D. Hartshorne and P. Cohen, respectively.

Protein sequencing and phosphopeptide analysis

Gel pieces containing purified protein were excised from Coomassie-stained SDS gels, loaded into pierced PCR tubes and sealed with DEC caps. The in-gel digestion process was performed automatically using DigestPro (Abimed). Peptide extracts were redissolved in 20 μ l 5–10% formic acid and 0.5 μ l were spotted onto a laser target, followed by analysis on a REFLEX MALDI time-of-flight mass spectrometer (Bruker Analytik). The peptide mass maps were recorded with mass resolution up to 15,000 and mass accuracy below 50 p.p.m. Data acquisition parameters, transfer and subsequent averaging of spectra, as well as further data processing, were carried out using the program LaserOne. For Edman sequencing, the peptides were purified by reversed phase HPLC with a 60-min linear gradient of 0.1% TFA and 80% acetonitrile. The separated peptides were sequenced on an Applied Biosystem 494 procise sequencer. For phosphopeptide analysis, the radioactive HPLC fractions were collected, dried and reconstituted in 50% methanol/5% ammonia for tandem mass spectrometric (MS/MS) analysis. All MS/MS experiments were carried out on an API III triple quadrupole mass spectrometer (PE Sciex). Aliquots (0.5 μ l) of reconstituted fractions were directly injected into the nanoelectrospray source and the phosphopeptides were detected by parent ion scans for PO₃⁻ (m/z 79) in negative ion mode²⁸. The mode of operation was then switched to positive for sequence determination of the identified phosphopeptides in product ion scan mode.

Cloning of IRAG cDNA from bovine tracheal muscle

Oligo-dT primed cDNA library from bovine tracheal muscle was constructed using the SuperScript Choice System (Gibco-BRL). mRNA (5 μ g) and primers (80 ng) were used for reverse transcription. Screening was performed with EST m78793 (Genbank) which was 91% identical to the nucleotide sequence of bovine IRAG. Primer extension library was constructed using specific oligo nucleotides 1,840–1,860 and probed with fragment nucleotides 1,489–1,828 of bIRAG.

Table 1 Modulation of bradykinin-induced Ca²⁺ transients by IRAG

Plasmids	AUC-ratio – cGMP	AUC-ratio + cGMP
cGKI β	101.5 $\pm$ 15.3 (12)	90.0 $\pm$ 6.2 (10)
bIRAG α	90.6 $\pm$ 9.4 (8)	87.9 $\pm$ 6.3 (7)
bIRAG α + cGKI β	87.7 $\pm$ 2.9 (17)	48.9 $\pm$ 5.9 (15)*†
bIRAG β	95.5 $\pm$ 10.7 (5)	96.2 $\pm$ 4.6 (8)
bIRAG β + cGKI β	103.1 $\pm$ 5.7 (16)	51.4 $\pm$ 9.8 (7)*

The ratio of the area under the curve (AUC) of second over the first transients is shown. The second transient was elicited in the absence (–cGMP) or presence (+cGMP) of 8-pCPT-cGMP. Values are $\bar{x} \pm$ s.e.m. (n = number of cells).

*P < 0.001 against cGKI β + cGMP.

†P < 0.001 against IRAG α + cGKI β – cGMP. For further details see Fig. 4.

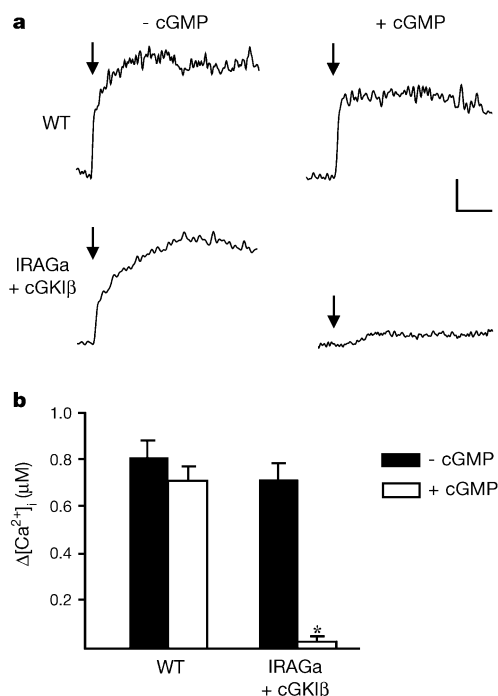


Figure 5 IRAG inhibits IP₃-evoked Ca²⁺ release. Both wild-type (WT) and bIRAG α /cGKI β -transfected (IRAG α + cGKI β) COS-7 cells were loaded with Fura-2 AM and were patched with pipettes containing 20 μ M IP₃. **a**, Calcium release was induced directly by IP₃ after perforating the membrane (arrow) of cells which were pre-incubated without (left panel) or with (right panel) 100 μ M 8-pCPT-cGMP for 5 min. The horizontal and vertical bars indicate 50 s and 200 nM Δ[Ca²⁺]_i, respectively. **b**, Summary of the experiments shown in **a**. Values are mean $\pm$ s.e.m. (n $\geq$ 6). Asterisk, p < 0.001 compared with IRAG α + cGKI β – cGMP.

Two-hybrid screen

The cDNA of cGKI β as a bait was inserted into yeast expression plasmid pEG202 in frame with the DNA-binding domain of LexA. A random-primed cDNA library of the rat intestinal longitudinal muscle was subcloned into yeast expression vector pJG4-5 containing the 'acid loop'-DNA activation domain. The library contained 1.2×10^6 independent clones and a total of 8.5×10^5 independent transformants were screened in EGY48 yeast cells. Interacting proteins were identified by colony selection on plates lacking leucine, tryptophan, histidine and uracil and confirmed by a β -galactosidase assay. The cDNA of positive clones was co-transformed into yeast either with bait, without bait or an unrelated bait to confirm the interaction.

GFP-assisted IRAG localization

Amino-terminal IRAG-GFP fusions were generated by inserting the appropriate fragments of IRAGa and IRAGb into expression vector pEGFP-N3 (Clontech). COS-7 cells were transfected transiently with pEGFP-IRAG using the calcium phosphate method. After 48 h of incubation, GFP fluorescence was visualised using a Axioplan microscope and filter set 09 (Zeiss).

Measurement of intracellular calcium

Cells were loaded with Fura-2 AM (5 μ M) before measuring $[Ca^{2+}]_i$ by the dual-wavelength microfluorescence technique⁷. For perfusion with inositol 1,4,5-trisphosphate (IP₃; Sigma) Fura-2 AM loaded COS-7 cells were patched in standard whole cell configuration. After establishing base line $[Ca^{2+}]_i$, the membrane was perforated and the recording was continued. Patch pipette solution contained in mM: 130 KCl, 10 NaCl, 2 MgCl₂, 10 HEPES pH 7.4, 1 MgATP, 0.1 K₂BAPTA (Sigma), 20 μ M IP₃ and 50 μ M Na₂Fura-2 (ICN)²⁹.

Received 2 December 1999; accepted 21 January 2000.

- Berridge, M. J., Bootman, M. D. & Lipp, P. Calcium—a life and death signal. *Nature* **395**, 645–648 (1998).
- Somlyo, A. P. & Somlyo, A. V. Signal transduction and regulation in smooth muscle. *Nature* **372**, 231–236 (1994).
- Missiaen, L. *et al.* The bell-shaped Ca^{2+} dependence of the inositol 1,4,5-trisphosphate-induced Ca^{2+} release is modulated by Ca^{2+} /calmodulin. *J. Biol. Chem.* **274**, 13748–13751 (1999).
- Michikawa, T. *et al.* Calmodulin mediates calcium-dependent inactivation of the cerebellar type I inositol 1,4,5-trisphosphate receptor. *Neuron* **23**, 799–808 (1999).
- Egenthaler, M., Lohmann, S. M., Walter, U. & Pilz, R. B. Signal transduction by cGMP-dependent protein kinases and their emerging roles in the regulation of cell adhesion and gene expression. *Rev. Physiol. Biochem. Pharmacol.* **135**, 173–209 (1999).
- Pfeifer, A. *et al.* Structure and function of cGMP-dependent protein kinases. *Rev. Physiol. Biochem. Pharmacol.* **135**, 105–149 (1999).
- Ruth, P. *et al.* Transfected cGMP-dependent protein kinase suppresses calcium transients by inhibition of inositol 1,4,5-trisphosphate production. *Proc. Natl Acad. Sci. USA* **90**, 2623–2627 (1993).
- Pfeifer, A. *et al.* Defective smooth muscle regulation in cGMP kinase I-deficient mice. *EMBO J.* **17**, 3045–3051 (1998).
- von der Leyen, H. *et al.* Gene therapy inhibiting neointimal vascular lesion: In vivo transfer of endothelial cell nitric oxide synthase gene. *Proc. Natl Acad. Sci. USA* **92**, 1137–1141 (1994).
- Chiche, J. D. *et al.* Adenovirus-mediated gene transfer of cGMP-dependent protein kinase increases the sensitivity of cultured vascular smooth muscle cells to the antiproliferative and pro-apoptotic effects of nitric oxide/cGMP. *J. Biol. Chem.* **273**, 34263–34267 (1998).
- Casnell, J. E. & Greengard, P. Guanosine 3':5'-cyclic monophosphate-dependent phosphorylation of endogenous substrate proteins in membranes of mammalian smooth muscle. *Proc. Natl Acad. Sci. USA* **71**, 1891–1895 (1974).
- Koga, T., Yoshida, Y., Cai, J. Q., Islam, M. O. & Imai, S. Purification and characterization of 240-kDa cGMP-dependent protein kinase substrate of vascular smooth muscle. Close resemblance to inositol 1,4,5-trisphosphate receptor. *J. Biol. Chem.* **269**, 11640–11647 (1994).
- Komalavilas, P. & Lincoln, T. M. Phosphorylation of the inositol 1,4,5-trisphosphate receptor by cyclic GMP-dependent protein kinase. *J. Biol. Chem.* **269**, 8701–8707 (1994).
- Haug, L. S., Jensen, V., Hvalby, O., Walaas, S. I. & Ostfold, A. C. Phosphorylation of the inositol 1,4,5-trisphosphate receptor by cyclic nucleotide-dependent kinases *in vitro* and in rat cerebellar slices *in situ*. *J. Biol. Chem.* **274**, 7467–7473 (1999).
- Rooney, T. A., Joseph, S. K., Queen, C. & Thomas, A. P. Cyclic GMP induces oscillatory calcium signals in rat hepatocytes. *J. Biol. Chem.* **271**, 19817–19825 (1996).
- Tertyshnikova, S., Yan, X. & Fein, A. cGMP inhibits IP₃-induced Ca^{2+} release in intact rat megakaryocytes via cGMP- and cAMP-dependent protein kinases. *J. Physiol. (Lond.)* **512**, 89–96 (1998).
- Baltensperger, K., Carafoli, E. & Chiesi, M. The Ca^{2+} -pumping ATPase and the major substrates of the cGMP-dependent protein kinase in smooth muscle sarcolemma are distinct entities. *Eur. J. Biochem.* **172**, 7–16 (1988).
- Fujita, S. *et al.* Characterization of major phosphoproteins in the cGMP-mediated protein phosphorylation system of vascular smooth muscle membranes. *J. Vasc. Res.* **36**, 299–310 (1999).
- Surks, H. K. *et al.* Regulation of myosin phosphatase by a specific interaction with cGMP-dependent protein kinase Ia. *Science* **286**, 1583–1587 (1999).
- Behrens, T. W. *et al.* Jaw1, a lymphoid-restricted membrane protein localized to the endoplasmic reticulum. *J. Immunol.* **153**, 682–690 (1994).
- Brennan, J. E., Xia, H., Chao, D. S., Black, S. M. & Bredt, D. S. Regulation of neuronal nitric oxide synthase through alternative transcripts. *Dev. Neurosci.* **19**, 224–231 (1997).
- Newton, C. L., Mignery, G. A. & Südhof, T. C. Co-expression in vertebrate tissues and cell lines of multiple inositol 1,4,5-trisphosphate (InsP₃) receptors with distinct affinities for InsP₃. *J. Biol. Chem.* **269**, 28613–28619 (1994).
- Lipp, P., Thomas, D., Berridge, M. J. & Bootman, M. D. Nuclear calcium signalling by individual cytoplasmic calcium puffs. *EMBO J.* **16**, 7166–7173 (1997).

- Keilbach, A., Ruth, P. & Hofmann, F. Detection of cGMP-dependent protein kinase isozymes by specific antibodies. *Eur. J. Biochem.* **208**, 467–473 (1992).
- Nguyen, T., Chin, W.-C. & Verdugo, P. Role of Ca^{2+} /K⁺ ion exchange in intracellular storage and release of Ca^{2+} . *Nature* **395**, 908–912 (1998).
- O'Rourke, F. *et al.* Ca^{2+} release by inositol 1,4,5-trisphosphate is blocked by the K⁺-channel blockers apamin and tetrapentylammonium ion, and a monoclonal antibody to a 63 kDa membrane protein: reversal of blockade by K⁺ ionophores nigericin and valinomycin and purification of the 63 kDa antibody-binding protein. *Biochem. J.* **300**, 673–683 (1994).
- Shaughnessy, J. D. Jr *et al.* Mrv1, a common MRV integration site in BXH2 myeloid leukemias, encodes a protein with homology to a lymphoid-restricted membrane protein Jaw1. *Oncogene* **18**, 2069–2084 (1999).
- Wilm, M., Neubauer, G. & Mann, M. Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **68**, 527–533 (1996).
- Broad, L. M., Armstrong, D. L. & Putney, J. W. Jr Role of inositol 1,4,5-trisphosphate receptor in Ca^{2+} feedback inhibition of calcium release-activated calcium current (I_{CR}). *J. Biol. Chem.* **274**, 32881–32888 (1999).

Acknowledgements

We thank S. Kamm and C. Wolf for technical assistance. This work was supported by Deutsche Forschungsgemeinschaft and Fond der Chemischen Industrie.

Correspondence and requests for materials should be addressed to F.H. (e-mail: pharma@ipt.med.tumuenchen.de). The Genbank accession numbers for IRAG cDNA sequences are AF195526 (IRAGa) and AF195527 (IRAGb).

hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response

Jong-Soo Lee, Kimberly M. Collins, Alexandra L. Brown, Chang-Hun Lee & Jay H. Chung

Laboratory of Molecular Hematology, NHLBI, NIH, Bethesda, Maryland 20892, USA

Mutations in the *BRCA1* (ref. 1) tumour suppressor gene are found in almost all of the families with inherited breast and ovarian cancers and about half of the families with only breast cancer^{2,3}. Although the biochemical function of *BRCA1* is not well understood, it is important for DNA damage repair^{4–7} and cell-cycle checkpoint^{8–10}. *BRCA1* exists in nuclear foci but is hyperphosphorylated and disperses after DNA damage^{11,12}. It is not known whether *BRCA1* phosphorylation and dispersion and its function in DNA damage response are related. In yeast the DNA damage response and the replication-block checkpoint are mediated partly through the Cds1 kinase family^{13–20}. Here we report that the human Cds1 kinase (hCds1/Chk2)^{21–23} regulates *BRCA1* function after DNA damage by phosphorylating serine 988 of *BRCA1*. We show that hCds1 and *BRCA1* interact and co-localize within discrete nuclear foci but separate after gamma irradiation. Phosphorylation of *BRCA1* at serine 988 is required for the release of *BRCA1* from hCds1. This phosphorylation is also important for the ability of *BRCA1* to restore survival after DNA damage in the *BRCA1*-mutated cell line HCC1937.

Table 1 hCds1 phosphorylates *BRCA1* in amino acid residues 785–1064 (GST-*BRCA1* 4)

BRCA1 fragment (amino acids)	Phosphorylation by hCds1
GST- <i>BRCA1</i> 1 (1–324)	–
GST- <i>BRCA1</i> 2 (260–553)	–
GST- <i>BRCA1</i> 3 (502–802)	–
GST- <i>BRCA1</i> 4 (758–1064)	+++
GST- <i>BRCA1</i> 5 (1005–1313)	–
GST- <i>BRCA1</i> 6 (1314–1863)	–

GST-*BRCA1* fusion proteins containing the indicated regions of *BRCA1* were used as substrate for hCds1 in kinase reactions. The level of phosphorylation by hCds1 for each region is shown: +++, strong; –, very weak or not detectable.

To determine whether human hCds1 directly phosphorylates BRCA1, six glutathione *S*-transferase (GST)–BRCA1 fusion proteins (numbered 1–6) containing overlapping BRCA1 fragments were used as substrates for recombinant hCds1 (Table 1). Of the six fusion proteins, hCds1 phosphorylated most strongly GST–BRCA1 4 (BRCA1 amino-acids 758–1,064). Consistent with this, gamma irradiation increased the ability of endogenous hCds1 to phosphorylate GST–BRCA1 4 almost sixfold (Fig. 1a).

By performing mass spectrometric analysis on hCds1-

phosphorylated GST–BRCA1 4, we determined that Ser988 of BRCA1 was the phosphorylated amino acid (Fig. 1b). Ser988 of human BRCA1 aligns with Ser971 of mouse and Ser987 of dog BRCA1. We performed a kinase reaction with recombinant hCds1 using as substrates GST–BRCA1 4 and GST–BRCA1 4 (S988A) in which Ser988 was changed to alanine. hCds1 did not phosphorylate GST–BRCA1 4 (S988A) confirming that hCds1 phosphorylated Ser988 in BRCA1 (Fig. 1c).

To determine whether Ser988 is also phosphorylated *in vivo*, we

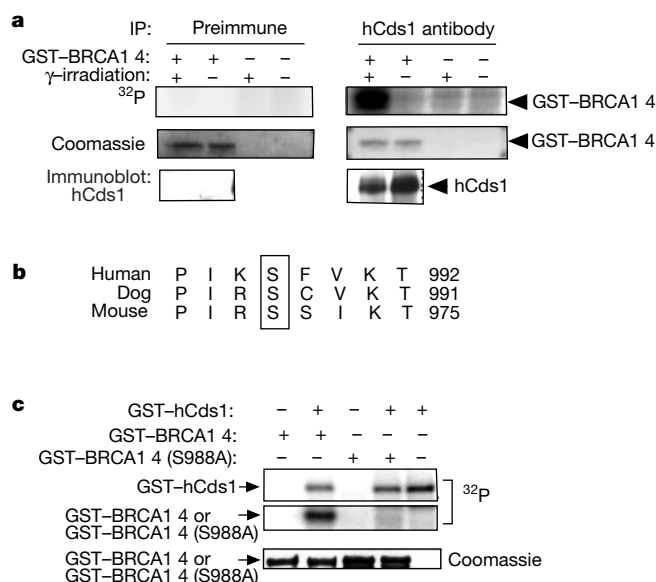


Figure 1 hCds1 phosphorylates Ser988 of BRCA1 *in vitro*. **a**, Endogenous hCds1 was immunoprecipitated from MCF-7 cells with hCds1-specific polyclonal antibody before (–) and 1 h after (+) gamma irradiation (10 Gy) and a kinase reaction was performed with GST–BRCA1 4 as substrate. Phosphorylation level (32 P) is shown at the top. **b**, Ser988 (boxed) of human BRCA1 is conserved in mammals. The position of the last threonine shown is indicated on the right. **c**, *In vitro* kinase reaction was performed with GST–hCds1 and substrates GST–BRCA1 4 or GST–BRCA1 4 (S988A).

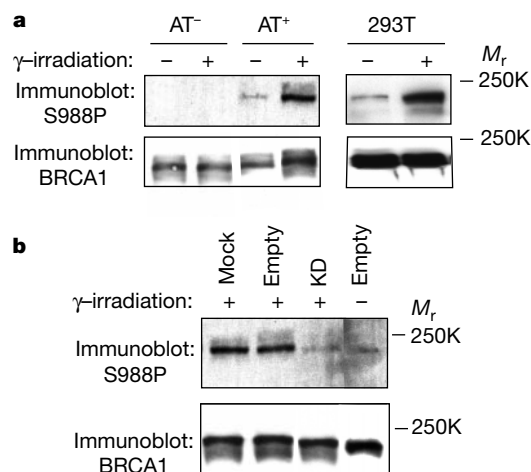


Figure 2 hCds1 phosphorylates Ser988 of human BRCA1 *in vivo*. **a**, Phosphorylation of Ser988 in AT[–], AT⁺ and 293T cells that were either untreated (–) or treated with 2.5 Gy of gamma irradiation (+) was visualized by immunoblotting with S988-P antibody. The total level of BRCA1 was visualized by immunoblotting with BRCA1 antibody (lower panel). **b**, Phosphorylation of Ser988 in 293T cells expressing kinase-dead hCds1 (KD) is shown. (The total level of BRCA1 is shown below).

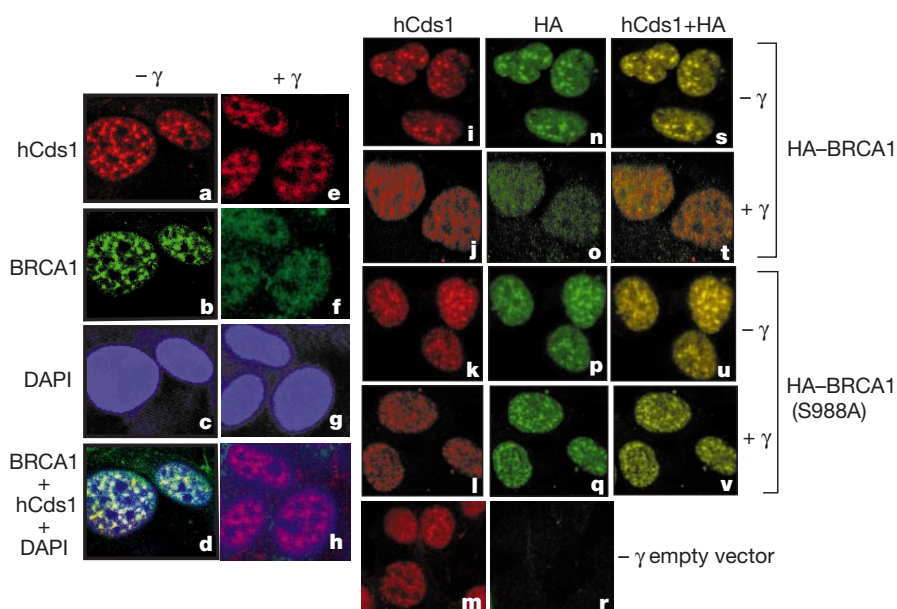


Figure 3 hCds1 and BRCA1 are localized to overlapping nuclear foci. **a–d**, Images of untreated (– γ) MCF-7 cells immunostained with anti-hCds1 antibody (**a**), anti-BRCA1 antibody (**b**) and DAPI (**c**). Overlap of these three images is shown in **d**. **e–h**, BRCA1 and hCds1 images 1 h after gamma irradiation (20 Gy; + γ). **i–v**, Images of endogenous

hCds1 (**i–m**) and transiently expressed HA–BRCA1 (**n**, **o**) and HA–BRCA1 (S988A) (**p**, **q**) in T24 cells detected with anti-HA antibody. Images of T24 cells transfected with an empty expression vector are shown in panels **j**, **o**, **t** and **l**, **q**, **v** were gamma irradiated. The overlap images are shown in panels **s–v**.

raised a polyclonal antibody (S988-P) that was specific for phosphorylated Ser 988. Using the S988-P antibody, we examined the phosphorylation status of BRCA1 in AT⁺ and AT⁻ cells that were gamma irradiated (Fig. 2a, left panel). In AT⁺ cells phosphorylation of Ser 988 increased significantly after gamma irradiation, whereas in AT⁻ cells phosphorylation of Ser 988 was not detectable. This is consistent with the fact that hCds1 activity is ATM-dependent at low doses of gamma irradiation^{21,23}. Gamma irradiation also increased phosphorylation of Ser 988 in 293T cells (Fig. 2a, right panel) but expression of kinase-dead hCds1 blocked it (Fig. 2b). Together, these results indicate that hCds1 may phosphorylate Ser 988 of BRCA1 *in vivo*. Notably, Ser 988 was phosphorylated at a low but significant level in untreated AT⁺ and 293T cells but not in AT⁻ cells. This implies that the ATM-hCds1-BRCA1 pathway is partially active in the basal state. In addition, the S988-P antibody detected a single band whereas the BRCA1 antibody detected a more diffuse band, indicating that DNA damage may cause phosphorylation of sites other than Ser 988.

Previous studies have shown that BRCA1 forms nuclear foci that disperse after DNA damage^{11,12}. Immunofluorescence staining of MCF-7 cells with polyclonal hCds1 antibody showed that hCds1 also exists in nuclear foci (Fig. 3a). Double staining of BRCA1 (green) and hCds1 (red) shows that hCds1 foci and BRCA1 foci co-localize (Fig. 3a–d, –γ). One hour after gamma irradiation, BRCA1 foci dispersed, resulting in a fainter pattern, and its co-localization with hCds1 foci diminished significantly (Fig. 3e–h, +γ). In contrast, hCds1 foci dispersed to a lesser degree, indicating that gamma irradiation may trigger the release of BRCA1 from hCds1 foci. To determine whether phosphorylation of Ser 988 is essential for dispersion of BRCA1, we examined the dispersion pattern of HA-BRCA1 and HA-BRCA1 (S988A) after gamma irradiation. Both HA-BRCA1 and HA-BRCA1 (S988A) co-localized with endogenous hCds1 in the absence of gamma irradiation (Fig. 3i, n, s and k, p, u, respectively). However, HA-BRCA1 dispersed strongly (Fig. 3o, t) whereas HA-BRCA1 (S988A) dispersed poorly

(Fig. 3q) and remained co-localized with hCds1 (Fig. 3v) after gamma irradiation. This suggests that phosphorylation of Ser 988 is essential for BRCA1 dispersion after gamma irradiation.

Thus, hCds1 and BRCA1 co-localize in nuclear foci but separate after gamma irradiation. We next examined with co-immunoprecipitation whether hCds1 and BRCA1 interact and, if so, whether gamma irradiation affects their interaction (Fig. 4a). Myc-tagged BRCA1 and V5-tagged hCds1 interacted before gamma irradiation but separated after gamma irradiation. Kinase-dead hCds1 and BRCA1 also interacted but did not separate after gamma irradiation, indicating that the kinase activity of hCds1 may be required for separation of the two proteins. Endogenous BRCA1 and hCds1 also interacted in MCF-7 cells but separated after gamma irradiation (Fig. 4b).

As the kinase activity of hCds1 was required for the release of BRCA1 (see Fig. 4a), we wanted to determine whether the phosphorylation of BRCA1 by hCds1 triggered its release. We therefore performed co-immunoprecipitation experiments with hCds1 and BRCA1 (S988A) (Fig. 4c). Unlike wild-type BRCA1, BRCA1 (S988A) was not released from hCds1 after gamma irradiation. This finding is consistent with decreased dispersion of BRCA1 (S988A) and its co-localization with hCds1 even after gamma irradiation (Fig. 3q, v). This result indicates that phosphorylation of BRCA1, and presumably the conformational change it induces in BRCA1, may be essential for its release from hCds1. Replacing Ser 988 with glutamic acid (S988E), therefore, might release BRCA1 from hCds1. BRCA1 (S988E) was released from hCds1 after gamma irradiation, but it still interacted with hCds1 before gamma irradiation, indicating that Ser 988 phosphorylation is required but not sufficient for their separation (Fig. 4c).

After DNA damage, hCds1 is autophosphorylated (data not shown). We examined the possibility that autophosphorylation of hCds1 may also be required for BRCA1-hCds1 separation. We repeated the co-immunoprecipitation study (Fig. 4c) with kinase-dead hCds1 (Fig. 4d). Unlike wild-type hCds1, kinase-dead hCds1 failed to separate from BRCA1 (S988E) even after gamma irradiation. This result indicates that modification of both hCds1 and BRCA1 is necessary for their separation.

HCC1937 breast cancer cells, which carry a homozygous mutation in BRCA1, are extremely sensitive to DNA damage. Exogenous expression of BRCA1 restores resistance to DNA damage in these cells²⁴. We compared BRCA1 (S988A) and BRCA1 (S988E) with wild-type BRCA1 (WT) for their ability to confer resistance to gamma irradiation using a published strategy²⁴. Expression of wild-type BRCA1 and BRCA1 (S988E), but not BRCA1 (S988A), con-

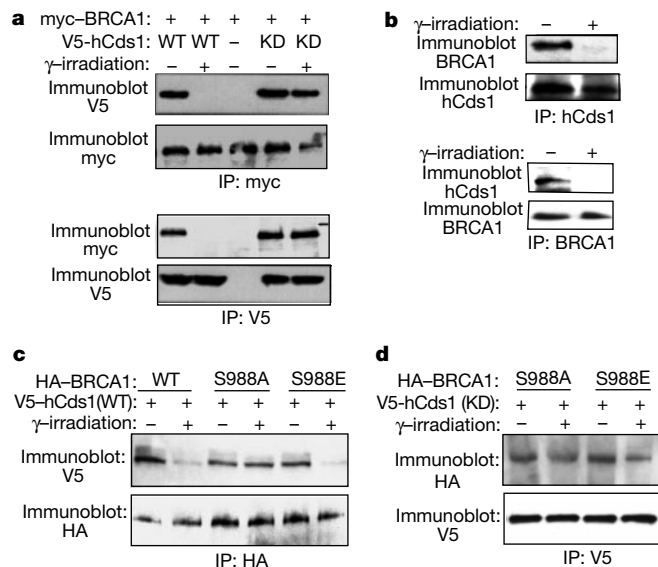


Figure 4 BRCA1-hCds1 interaction is sensitive to DNA damage. **a**, Interaction between myc-BRCA1 and V5-hCds1 visualized by co-immunoprecipitation before (–) or 1–2 h after (+) gamma irradiation (10 Gy). The antibody used for immunoprecipitation is indicated below (IP). **b**, Interaction between endogenous BRCA1 and hCds1 in MCF-7 cells visualized by co-immunoprecipitation using hCds1- (top) or BRCA1- (bottom) specific antibodies before (–) and after (+) gamma irradiation. **c**, Interaction between V5-hCds1 (WT) and HA-BRCA1 (WT, S988A or S988E) visualized by co-immunoprecipitation. **d**, Interaction between V5-hCds1 (KD) and HA-BRCA1 (S988A or S988E) visualized as above.

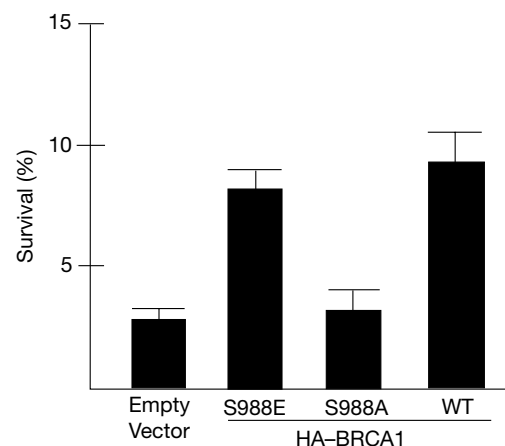


Figure 5 Phosphorylation of Ser 988 is important for BRCA1 function. Survival of HCC1937 cells expressing HA-BRCA1 (WT, S988A and S988E) eight days after gamma irradiation is shown.

ferred resistance to gamma irradiation in HCC1937 cells (Fig. 5). This result is consistent with the idea that the phosphorylation of Ser988 by hCds1 is important for BRCA1 function in the DNA damage response.

Our results indicate that the DNA damage pathway involving evolutionarily conserved hCds1 regulates BRCA1 by direct phosphorylation which controls its dispersion and function after DNA damage. However, BRCA1 phosphorylation may be more complicated. For example, serines 1,423 and 1,524 of BRCA1 can be phosphorylated by ATM after a high dose of gamma radiation²⁵. In addition, Ser 1,497 appears to be phosphorylated by Cdk2 during the G1/S phase of the cell cycle²⁶. Phosphorylation of the different serine residues is likely to have different effects on BRCA1 function: a challenge for the future is to define these effects. □

Methods

Plasmid construction

The S988A and S988E mutations in HA-tagged BRCA1²⁷ were created by using the MORPH site-specific plasmid DNA mutagenesis kit (Eppendorf-5 Prime) with oligonucleotides 5'-CCACTTTTCCCATCAAGGCCCTTGTAAACTAAATG-3' and 5'-CCACTTTTCCCATCAAGGAATTTGTAAACTAAATG-3', respectively. The S988A mutation in GST-BRCA1 4 was created by using the QuikChange site-directed mutagenesis kit (Stratagene) and oligonucleotides 5'-CCACTTTTCCCATCAAGGCC-TTGTGTAAACTAAATG-3' and 5'-CATTTAGTTTAAACAAAGGCCCTGATGGG-AAGAGTGG-3'. The mutations were verified by sequencing.

Transfection

Transfections were performed with the Effectene kit (Qiagen) using the manufacturer's recommendations. For co-immunoprecipitation experiments, 10⁷ COS-7 or 293T cells were transfected with 2 µg of V5-hCds1 and 2 µg of either HA-BRCA1 or myc-BRCA1 expression vectors by using Effectene.

Antibodies

To detect endogenous proteins, we used affinity purified rabbit anti-hCds1 (ref. 23) and anti-BRCA1 antibodies (monoclonal; Ab-3, Calbiochem; or rabbit polyclonal, BRCA1 1847-1863, Pharmingen). To detect myc-BRCA1 and V5-hCds1, anti-myc-HRP (Invitrogen) and anti-V5-HRP (Invitrogen) antibodies were used. To detect HA-BRCA1, we used anti-HA antibody (HA.11, Babco or 3F1, Boehringer Mannheim).

Generation of S988-P antibody and immunoblotting

Synthetic phosphopeptide S988-P (CRIPPLFIKIS(988)FVKTK), in which Ser988 was phosphorylated, was conjugated by its N-terminal cysteine to keyhole limpet haemocyanin (KLH, Fluka) using the heterobifunctional crosslinker, *N*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (sulpho-MBS, Pierce). New Zealand white rabbits were immunized with 1 mg KLH conjugate suspended in complete adjuvant and boosted with 0.3 mg conjugate in incomplete adjuvant after 4 and 7 weeks. S988-P antibody was purified by using an affinity column containing the phosphopeptide which was prepared using activated CH-Sepharose 4B (Pharmacia) according to the manufacturer's instructions. The specificity of S988-P antibody was demonstrated by showing that it does not recognize GST-BRCA1 4 (S988A) mutant. To detect *in vivo* phosphorylation of Ser988, the AT⁺ (+ or -)²⁸ cells (2 × 10⁷) or the 293T cells (0.5–0.8 × 10⁷) were first treated with gamma irradiation (2.5 Gy). One hour later, the cells were lysed and BRCA1 was immunoprecipitated with an anti-BRCA1 antibody (Ab-3, Calbiochem). Immunoprecipitated BRCA1 was electrophoresed on denaturing PAGE and then transferred to a nitrocellulose filter. Immunoblotting was performed with S988-P antibody (1:200 dilution) that had been preadsorbed with 1 mg of 293T cell extract in 1 ml of PBS containing 0.1% Tween 20 for 1 h at room temperature. To determine the total amount of BRCA1 in each sample, immunoblotting was also performed with anti-BRCA1 antibody (Pharmingen).

In vitro kinase assay

GST-BRCA1 fusion proteins (numbered 1–6) were prepared as described in ref. 27. *In vitro* kinase assays were performed in a 15 µl total volume with 1 µg of recombinant GST-hCds1 (WT or KD)²³ and 2 µg of GST-BRCA1 fusion protein (1–6) substrate in a reaction buffer containing 10 mM HEPES (pH 7.5), 75 mM KCl, 5 mM MgCl₂, 0.5 mM EDTA, 1 mM dithiothreitol, 100 µM ATP and 5 µCi of [γ -³²P]ATP. The mixture was incubated at 30 °C for 15 min and electrophoresed in a denaturing polyacrylamide gel. The effect of gamma irradiation on hCds1 activity was examined by immunoprecipitating hCds1 with 1 µg of preimmune serum or 1 µg of affinity-purified antibody specific for hCds1 (ref. 23) from MCF-7 cell lysate (10⁷ cells) prepared either before or 1 h after gamma irradiation (10 Gy). The kinase reactions were performed as described above using GST-BRCA1 4 as substrate.

Immunofluorescence staining and confocal microscopy

Immunofluorescence staining of BRCA1 and hCds1 was done essentially as described^{24,29} except that 2 µg ml⁻¹ of affinity-purified polyclonal rabbit antibody against hCds1 (ref. 23)

was used where indicated. Confocal laser scanning microscopy was performed on the immunofluorescently stained cells with a Noran Odyssey real-time laser confocal microscope equipped with a Nikon Diaphot inverted microscope.

Conferring gamma irradiation resistance on HCC1937 cells by using BRCA1

The function of BRCA1 in conferring gamma irradiation resistance to HCC1937 cells was examined as described²⁴ with the following modifications. HCC1937 (10⁵) cells were transfected with expression vectors for HA-BRCA1 (WT), HA-BRCA1 (S988A), HA-BRCA1 (S988E) or empty vector using Effectene (Qiagen). Forty-eight hours later, the cells were either untreated or gamma irradiated (0.14 Gy) for each transfection.

Received 7 October 1999; accepted 20 January 2000.

- Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
- Easton, D. F., Bishop, D. T., Ford, D. & Crookford, G. P. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **52**, 678–701 (1993).
- Feunteun, J. & Lenoir, G. M. BRCA1, a gene involved in inherited predisposition to breast and ovarian cancer. *Biochim. Biophys. Acta* **1242**, 177–180 (1996).
- Gowen, L. C., Avrutskaya, A. V., Latour, A. M., Koller, B. H. & Leadon, S. A. BRCA1 required for transcription-coupled repair of oxidative DNA damage. *Science* **281**, 1009–1012 (1998).
- Shen, S. X. *et al.* A targeted disruption of the murine Brca1 gene causes gamma-irradiation hypersensitivity and genetic instability. *Oncogene* **17**, 3115–3124 (1998).
- Abbott, D. W. *et al.* BRCA1 expression restores radiation resistance in BRCA1-defective cancer cells through enhancement of transcription-coupled DNA repair. *J. Biol. Chem.* **274**, 18808–18812 (1999).
- Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. Brca1 controls homology-directed DNA repair. *Mol. Cell* **4**, 511–518 (1999).
- Somasundaram, K. *et al.* Arrest of the cell cycle by the tumour-suppressor BRCA1 requires the CDK-inhibitor p21WAF1/CIP1. *Nature* **389**, 187–190 (1997).
- Larson, J. S., Tonkinson, J. L. & Lai, M. T. A BRCA1 mutant alters G2-M cell cycle control in human mammary epithelial cells. *Cancer Res.* **57**, 3351–3355 (1997).
- Xu, X. *et al.* Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. *Mol. Cell* **3**, 389–395 (1999).
- Scully, R. *et al.* Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage. *Cell* **90**, 425–435 (1997).
- Thomas, J. E., Smith, M., Tonkinson, J. L., Rubinfeld, B. & Polakis, P. Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. *Cell Growth Differ.* **8**, 801–809 (1997).
- Zheng, P. *et al.* SPK1 is an essential S-phase-specific gene of *Saccharomyces cerevisiae* that encodes a nuclear serine/threonine/tyrosine kinase. *Mol. Cell. Biol.* **13**, 5829–5842 (1993).
- Allen, J. B., Zhou, Z., Siede, W., Friedberg, E. C. & Elledge, S. J. The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev.* **8**, 2401–2415 (1994).
- Murakami, H. & Okayama, H. A kinase from fission yeast responsible for blocking mitosis in S phase. *Nature* **374**, 817–819 (1995).
- Sanchez, Y. *et al.* Regulation of RAD53 by the ATM-like kinases MEC1 and TEL1 in yeast cell cycle checkpoint pathways. *Science* **271**, 357–360 (1996).
- Sidorova, J. M. & Breeden, L. L. Rad53-dependent phosphorylation of Swi6 and down-regulation of CLN1 and CLN2 transcription occur in response to DNA damage in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 3032–3045 (1997).
- Lindsay, H. D. *et al.* S-phase-specific activation of Cds1 kinase defines a subpathway of the checkpoint response in *Schizosaccharomyces pombe*. *Genes Dev.* **12**, 382–395 (1998).
- Boddy, M. N., Furnari, B., Mondesert, O. & Russell, P. Replication checkpoint enforced by kinases Cds1 and Chk1. *Nature* **395**, 507–510 (1998).
- Zeng, Y. *et al.* Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1. *Nature* **395**, 507–510 (1998).
- Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
- Blasina, A. *et al.* A human homologue of the checkpoint kinase Cds1 directly inhibits Cdc25 phosphatase. *Curr. Biol.* **9**, 1–10 (1999).
- Brown, A. L. *et al.* A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage. *Proc. Natl. Acad. Sci. USA* **96**, 3745–3750 (1999).
- Zhong, Q. *et al.* Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science* **285**, 747–750 (1999).
- Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* **286**, 1162–1166 (1999).
- Ruffner, H., Jiang, W., Craig, A. G., Hunter, T. & Verma, I. M. BRCA1 is phosphorylated at serine 1497 *in vivo* at a cyclin-dependent kinase 2-phosphorylation site. *Mol. Cell. Biol.* **19**, 4843–4854 (1999).
- Scully, R. *et al.* Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* **88**, 265–275 (1997).
- Ziv, Y. *et al.* Recombinant ATM protein complements the cellular A-T phenotype. *Oncogene* **15**, 159–167 (1997).
- Wilson, C. A. *et al.* Localization of human BRCA1 and its loss in high-grade, non-inherited breast carcinomas. *Nature Genet.* **21**, 236–240 (1999).

Acknowledgements

We thank D. Livingston and R. Scully for the BRCA1 plasmids; Y. Shiloh for the AT cells; M. Kirby for flow cytometry; and Y. Xu and C. Combs for help with confocal microscopy. We also thank W. S. Lane, D. P. Kirby and K. A. Pierce of the Harvard Microchemistry facility for expertise in tandem mass spectrometry peptide sequencing.

Correspondence and requests for materials should be addressed to J.H.C. (e-mail: jhchung@helix.nih.gov).

Ligand binding and conformational motions in myoglobin

Andreas Ostermann*, Robert Waschipyk†, Fritz G. Parak* & G. Ulrich Nienhaus‡,§

*Fakultät für Physik E17, Technische Universität München, D-85747 Garching, Germany

†Department of Biophysics, University of Ulm, D-89069 Ulm, Germany

‡Department of Physics, University of Illinois, Urbana, Illinois 61801-3080, USA

Myoglobin, a small globular haem protein that binds gaseous ligands such as O₂, CO and NO reversibly at the haem iron, serves as a model for studying structural and dynamic aspects of protein reactions. Time-resolved spectroscopic measurements after photodissociation of the ligand revealed a complex ligand-binding reaction with multiple kinetic intermediates, resulting from protein relaxation and movements of the ligand within the protein^{1–3}. To observe the structural changes induced by ligand dissociation, we have carried out X-ray crystallographic investigations of carbon monoxy-myoglobin (MbCO mutant L29W) crystals illuminated below and above 180 K, complemented by time-resolved infrared spectroscopy of CO rebinding. Here we show that below 180 K photodissociated ligands migrate to specific sites within an internal cavity—the distal haem pocket—of an essentially immobilized, frozen protein, from where they subsequently rebind by thermally activated barrier crossing. Upon photodissociation above 180 K, ligands escape from the distal pocket, aided by protein fluctuations that transiently open exit channels. We recover most of the ligands in a cavity on the opposite side of the haem group.

After dissociation from the haem iron at physiological temperatures, ligands either rebind internally (geminate) from within the distal haem pocket or escape into the solvent, assisted by protein fluctuations that transiently open exit channels^{1–5}. The extent of geminate recombination depends on the intrinsic reactivity of the ligand with the haem iron and its ability to diffuse away from the active site, as shown by X-ray crystallography, time-resolved spectroscopy and molecular dynamics simulations that have been carried out on a large number of mutant myoglobins^{3,6,7}. Lowering the temperature slows the protein motions that permit ligand escape, and the geminate fraction increases^{1,2}. At ~180 K, a dynamical transition occurs in which large-scale protein fluctuations become arrested^{8–11}, so that ligands remain in the haem pocket after dissociation.

Crystal structure analyses of the myoglobin (Mb) photoproduct intermediate, Mb*CO, at liquid-helium temperatures (20–40 K) have revealed the structural changes that occur upon photodissociation below the transition temperature^{12–14}: the CO moves to a docking site above haem pyrrole C, the haem group shifts by ~0.2 Å towards the distal side, and the iron shifts by ~0.3 Å out of the mean haem plane as a result of its spin transition and coordination change. Only small changes occur in the atomic positions of the frozen-in polypeptide moiety. This low-temperature photoproduct (with inhibited relaxation) is essentially identical to the short-time photoproduct (before protein relaxation) at room temperature^{15–17}. Its docking site is functionally important because it prevents fast recombination before the ligand can leave the haem pocket.

Protein relaxation and ligand migration can be observed with X-ray diffraction after photolysis of MbCO crystals at ~180 K. These experiments, however, incur the problem that recombination in native MbCO is so fast that the photoproduct yield is only ~10% under steady-state illumination using light intensities that do not damage the crystals¹⁸. To obtain 100% photoproduct, we chose the

Table 1 Data collection, refinement statistics and active site geometry

	Crystal I		Crystal II	
	105K dark	Illumination at T > 180 K	Illumination at T < 180 K	5h at T = 180 K after illumination
Data collection				
Space group	P6		P6	
Unit cell (Å)	a = b = 90.46, c = 45.26		a = b = 90.43, c = 45.24	
Crystal size (mm ³)	0.4 × 0.6 × 0.2		0.3 × 0.5 × 0.15	
Resolution (Å)	1.5	1.55	1.7	1.85
Measured reflections	175,770	190,950	140,474	108,921
$\langle I \rangle / \langle \sigma(I) \rangle$ *	14.6(5.1)	14.0(4.2)	14.0(4.3)	15.8(5.5)
Unique reflections	32,053	29,266	22,384	16,415
Completeness (%)	94.5(82.3)	95.1(84.1)	95.8(86.4)	90.3(76.7)
Multiplicity*	5.5(3.1)	6.5(3.1)	6.3(2.7)	6.6(2.3)
R _{merge} (%)†	4.3	4.6	5.0	4.6
Temperature (K)	105	105	105	105
Refinement				
Resolution range (Å)	7.0–1.5	7.0–1.55	7.0–1.7	7.0–1.85
Number of reflections	31,820	29,030	22,148	16,415
Working set/test set	28,630/3,190	26,130/2,900	19,958/2,190	15,129/1,286
R _{cryst} (%)‡	18.9	19.3	18.4	18.6
R _{free} (%)§	21.5	21.8	20.9	21.4
R.m.s. deviations bond length (Å)	0.007	0.007	0.007	0.007
R.m.s. deviations bond angles (°)	1.1	1.1	1.1	1.1
R.m.s. coordinate error (Å)	0.14	0.15	0.16	0.16
Geometry				
Fe–CO(Å)	1.85	–	–	1.87
Bond angle (°)¶	166	–	–	170
Tilt angle (°)#	8.6	–	–	8.9
Fe–(haem plane) (Å)²	0.10	0.29	0.31	0.14
Fe–(N ₄ plane) (Å)**	0.01	0.19	0.21	0.03
Fe–N _{H2} H93 (Å)	2.14	2.16	2.17	2.14

* Values in parentheses refer to the highest resolution shell.

† $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i - \langle I \rangle| / (\sum_{hkl} \sum_i I_i)$, where I_i denotes the scaled intensity of the i th symmetry-related observation of reflection hkl ; $\langle I \rangle = \Sigma_i I_i$.

‡ $R_{\text{cryst}} = \sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} |F_o(hkl)|$, where $F_o(hkl)$ and $F_c(hkl)$ are the observed and calculated structure factor amplitudes for the reflection hkl .

§ R_{free} is R_{cryst} calculated with a test set of observations that were excluded from the refinement calculations.

|| As estimated by a σ_A plot³¹.

¶ Defined by the iron and the C and O atoms of the haem-bound ligand.

Deviation of the Fe–C bond from the haem normal.

² Distance between Fe and the best-fit plane through the 20 central C atoms of the haem.

** Distance between Fe and the best-fit plane through the four haem nitrogen atoms.

mutant MbCO L29W (leucine 29 → tryptophan), which has a 130-fold smaller association rate coefficient than the native protein at room temperature¹⁹.

We took X-ray data sets of both the ligand-bound and the photoproduct state of L29W MbCO crystals at 105 K after illumination at $T < 180$ K. Data collection and refinement statistics are compiled in Table 1 (columns 1 and 3), together with geometrical details of the active site. Electron density maps and structural models are shown in Fig. 1. No electron density is seen at the ligand-binding site of the photoproduct, implying complete photodissociation has occurred. Neither do we observe electron density at the previously found CO location above pyrrole C (refs 12, 14). Instead, the ligand is located in a cavity in the back of the distal pocket. This location is known as the 'xenon 4' site; it is one of four internal cavities in myoglobin that bind xenon with high affinity²⁰. The displacements of the haem group and the iron atom are essentially identical to those observed in the helium-temperature photoproduct structures. Small shifts occur throughout the back-

bone, in particular in the E, F, and G helices, as seen from the r.m.s. differences between the average positions of the backbone atoms of the ligand-bound and photoproduct structures in Fig. 2. These light-induced changes are entirely reversible, as shown by collection of a data set at 105 K (Table 1, column 4) after photodissociation below 180 K and subsequent recombination at 180 K for 5 hours. Within experimental error, the structure thus obtained is identical to the one without prior illumination.

To determine the structure of the relaxed conformation, an L29W MbCO crystal was illuminated above 180 K before data collection at 105 K (Table 1, column 2). The electron density map and structural model in Fig. 3 show much larger changes than with illumination below 180 K. The backbone displacements, included in Fig. 2, reveal large shifts in the region of helices B, C and E, and still substantial displacements in the region of helices F and G. Most obvious in Fig. 3 are the rearrangements that occur in the vicinity of W29, including the backbone. The H64 imidazole side chain moves deeper into the haem pocket, and the indole group of W29 rotates by about 45° into

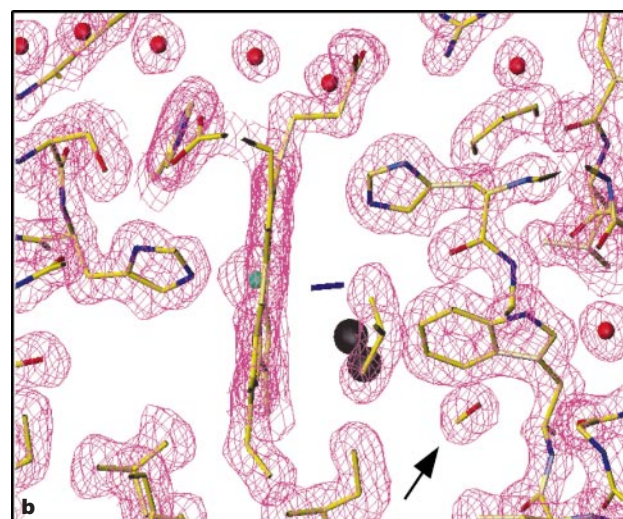
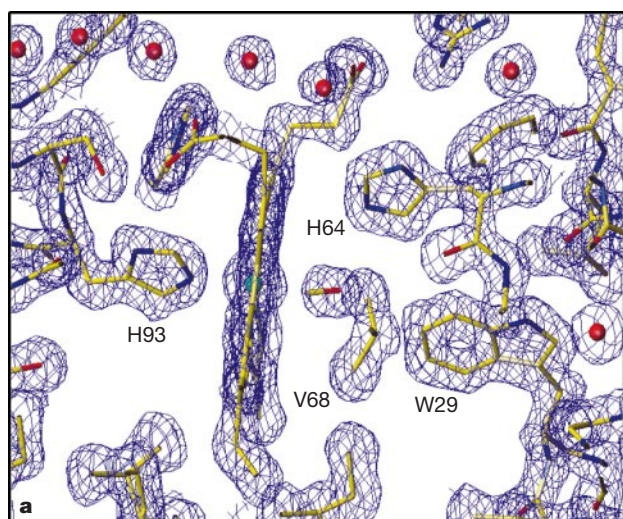


Figure 1 $(2F_o - F_c)\exp(i\alpha_c)$ electron density maps of the active site of L29W MbCO at 105 K, contoured at 1.5σ (standard deviation). F_o and F_c denote the observed and calculated structure-factor amplitudes and α_c the phases calculated from the refined model. In the stick models, C atoms are shown in yellow, N atoms in blue and O atoms in red. **a**, MbCO dark structure, with the iron in the haem plane and the CO bound in a single orientation. **b**, Mb*CO, photodissociated at temperatures below 180 K. The location of the CO is marked by an arrow; its orientation was arbitrarily assigned. A blue stick marks the bound CO position and the black dumbbell the photodissociated CO in the B₂ photoproduct state of native MbCO at 36 K (ref. 14).

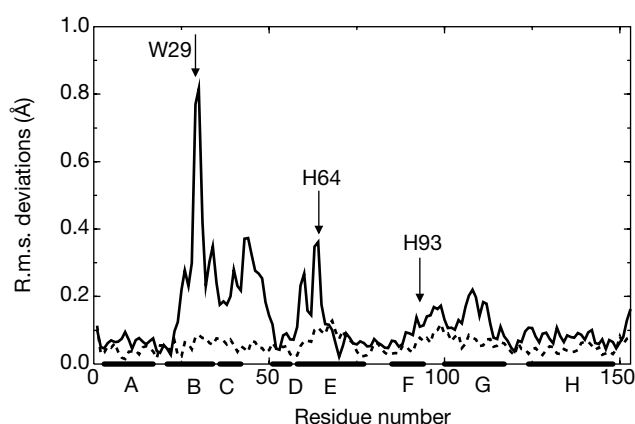


Figure 2 Root-mean-squared differences of the amino-acid positions (backbone average, N-Cα-C) between MbCO and the two structures created by photodissociation below 180 K (dashed line) and above 180 K (solid line), as a function of the sequence number. The eight helices of myoglobin, labelled A to H, are marked on the abscissa.

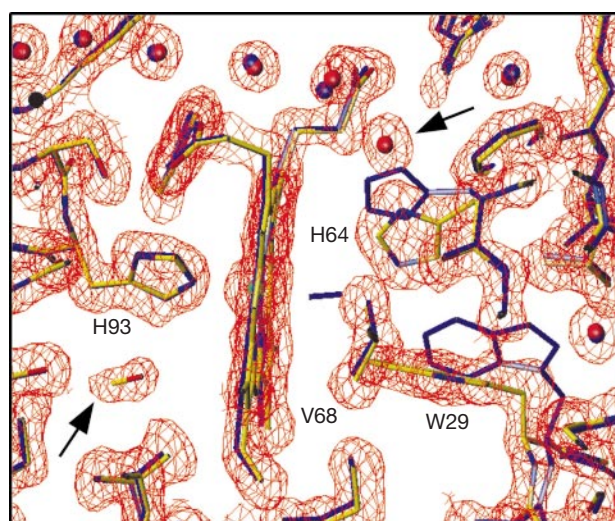


Figure 3 $(2F_o - F_c)\exp(i\alpha_c)$ electron density map, contoured at 1.5σ , of the active site of L29W MbCO at 105 K after extended illumination above 180 K and stick model with colour code as in Fig. 1. Also included is the model of the MbCO dark structure (Fig. 1a) in blue. Two new electron density features appearing in this structure are marked by arrows. The one on top of the H64 side chain is most probably due to a water molecule, whereas the one in the proximal pocket is attributed to a CO molecule.

a position close to the one found in the room temperature structure of L29W MbO₂ (ref. 21). Two new features appear in the electron density map, marked by arrows. We attribute the spherical one above the H64 imidazole to a water molecule because it is in the proper hydrogen-bonding distance to the N_{δ1} nitrogen of H64. Moreover, the additional water molecule and the shift of the imidazole are also seen when comparing ligated and deoxy forms of various mutant myoglobins⁶. The second, elongated feature on the proximal side shows up in the 'xenon 1' site²⁰. This cavity was the first xenon-binding site discovered²² and is the one with the highest affinity. Because no electron density appears at this location in either MbCO or deoxy-Mb, we have assigned this feature to a CO molecule; refinement yields a partial occupancy of 62%.

Molecular dynamics simulations²³ and kinetic studies of the effects of xenon on ligand binding⁷ suggest that the 'xenon 1' site is a physiologically relevant site on the migration path (or one of the paths^{5,23}) of the ligand between haem iron and solvent. To examine ligand migration from this location, we collected three more X-ray data sets at 105 K in succession (these data will be published elsewhere), interleaved by heat-cool cycles to ~180 K to enhance recombination, so that each structure had ~20% more rebound ligands than the previous one. The fractional population of CO in the proximal site scales, within error, with the fraction of ligands missing from the haem iron. This behaviour is expected if the CO molecules in the proximal cavity are in equilibrium with the other free CO molecules in the crystal at ~180 K.

We used time-resolved infrared spectroscopy to relate the observed structural changes to the ligand-binding function. In Fig. 4a and b, the fraction of proteins that have not yet rebound a ligand at time t after the photolysing laser flash, $N(t)$, is plotted versus t in double-logarithmic plots. The kinetics are non-exponential because the protein molecules exist in different conformational substates with different enthalpy barriers for rebinding^{24,25}. Up to 50 K, recombination in the L29W mutant is faster than in native MbCO. Starting at 60 K, a second kinetic species with markedly slower rebinding appears. Its proportion changes with temperature, implying that the two species can interconvert. The kinetics can be described quantitatively by the three-well model with two photoproduct states, B₂, B₃, and the

bound state A, depicted in Fig. 4c. The relevant kinetic equations are given in Methods. The solid lines in Fig. 4a and b are fits of this model to the data, the best-fit parameters are included in the inset of Fig. 4c. The fast-rebinding state is denoted by B₂ because it absorbs at 2,123 cm⁻¹, close to the 2,121 cm⁻¹ of B₂ in native Mb*CO (ref. 26), which has been suggested to differ from B₁ by the orientation of the CO ligand in the docking site on top of haem pyrrole C (ref. 15). The similarity of the CO stretch band and the availability of the position above pyrrole C suggest a similar ligand location in L29W MbCO. The photoproduct in Fig. 1b is identified with the slowly rebinding state B₃. Whereas B₃ can be populated by thermal activation in L29W MbCO, slowly rebinding photoproduct states of native MbCO can only be populated appreciably by extended illumination ('pumping')²⁷.

Above 180 K, yet another kinetic component appears at the longest times which is attributed to bimolecular ligand recombination after the photodissociated ligand has left the protein^{1,2}. Its fraction increases rapidly and dominates above 200 K (Fig. 4b). Between 200 and 300 K, the association rate coefficient of solvent rebinding increases according to the Arrhenius law, with an activation enthalpy of 36 kJ mol⁻¹. This suggests that no new physical phenomena occur in this temperature range and that the processes observed around 200 K are immediately relevant to ligand binding at physiological temperatures. Figure 4d shows Fourier transform infrared difference spectra of the photodissociated CO in states B₂ and B₃, and after leaving the distal pocket, denoted here by C.

Our structural and kinetic investigations of protein relaxation and CO migration in myoglobin after photodissociation below and above the dynamic transition temperature emphasize the crucial interplay between protein dynamics and function and support the following picture of the physiologically important ligand escape process: within picoseconds after dissociation from the haem iron, ligands either rebound or move to specific docking sites in the distal haem pocket. Recombination from these sites is slowed substantially, and thus a large fraction of ligands avoids rebinding long enough that large-scale protein fluctuations on nanosecond to microsecond timescales open exit channels through which ligands migrate into other internal cavities and finally escape from the protein. □

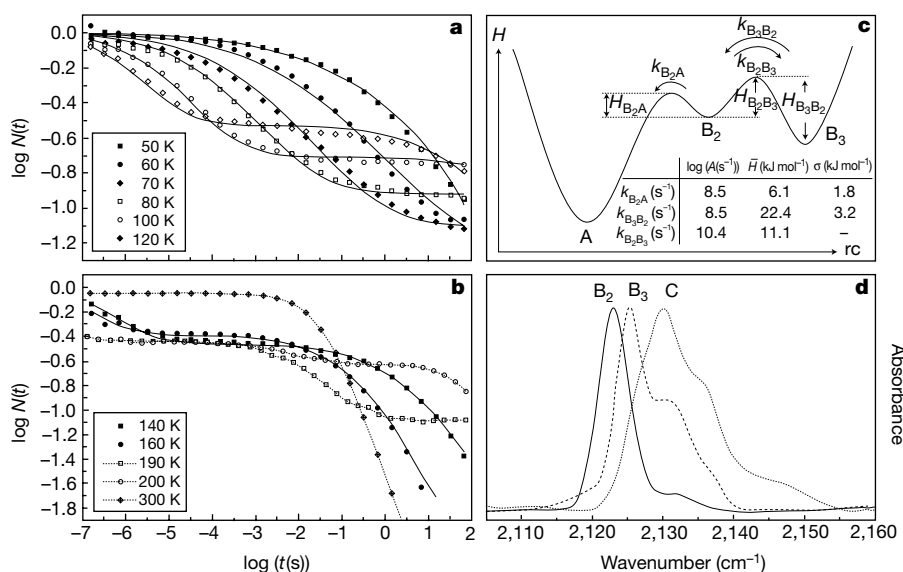


Figure 4 Flash photolysis kinetics of L29W MbCO below 180 K, measured in the infrared band of the bound CO at 1,945 cm⁻¹. **a, b**, Experimental data (symbols) for 50 K ≤ T ≤ 120 K (**a**) and 140 K ≤ T ≤ 300 K (**b**), including fits (solid lines) using the model sketched in **c**. **c**, The three-well model that describes the low-temperature kinetics, with bound state A and two photoproduct states, B₂, B₃. Details of the kinetic analysis are given in

Methods. Arrhenius parameters of the microscopic rate coefficients are compiled in the inset. **d**, Fourier-transform infrared difference spectra of the photodissociated CO in state B₂, taken after photolysis at 12 K, state B₃, taken at 12 K after cooling under illumination from 160 K, and state C, taken at 140 K after photolysis above 180 K.

Methods

Sample preparation

Mutant sperm whale myoglobin L29W was expressed in *Escherichia coli* and purified as described²⁸. For time-resolved spectroscopy, L29W MbCO was dissolved at a concentration of 20 mM in cryosolvent (75% glycerol, 25% potassium phosphate buffer (v/v) pH 7.5). For X-ray crystallography, crystals (space group *P6*) were grown in 2.5 M (NH₄)₂SO₄ solution, buffered with Tris-HCl to pH 8.5. Trehalose (300 mg ml⁻¹) was added to the mother liquor to avoid crystal damage upon cooling. Crystals were mounted using the loop technique and shock-frozen in a liquid nitrogen cryostream.

Infrared spectroscopy

Kinetics were measured by monitoring absorption changes in the infrared stretch band of haem-bound CO at 1,945 cm⁻¹. The protein sample was photolysed with 6 ns wide pulses (30 mJ) from a frequency-doubled Nd-YAG laser (532 nm). Rebinding was monitored with infrared light from a lead-salt laser diode (Laser Photonics), which was focused on the sample and the detector with off-axis paraboloid mirrors. The signal from the photo-voltaic InSb infrared detector (SAT Infrared Detectors) was digitized with a logarithmic time-base digitizer from 1 μs to 100 s and from 20 ns to 400 μs with a digital storage oscilloscope (Tektronix TDS 430A). A closed-cycle helium refrigerator (Helix Technology Corporation) cooled the laser diode and the sample. Steady-state Fourier-transform infrared difference spectra were taken with a Bruker IFS 66v/S spectrometer.

Kinetic model

The ligand-binding kinetics can be described by a three-well model, with the bound state A and two photoproduct states, B₂ and B₃, as sketched in Fig. 4c. Immediately after photodissociation, state B₂ is populated. The subsequent kinetics are governed by the three rate coefficients k_{B_2A} , $k_{B_2B_3}$, $k_{B_3B_2}$ (k_{AB_2} can be neglected here) and given by

$$N(t) = N_f \exp[-(k_{B_2A}/N_f)t] + N_s \exp[-(k_{B_3B_2} \times N_f)t]$$

where N_f and N_s denote the fast and the slowly rebinding fractions, respectively. Their population ratio is given by $N_f/N_s = k_{B_2A}/k_{B_3B_2}$; $N_f + N_s = 1$. To describe the non-exponentiality of the kinetics, we introduced gaussian distributions of enthalpy barriers for H_{B_2A} and $H_{B_3B_2}$, $g(H_i) = (2\pi\sigma_i^2)^{-1/2} \exp[-(H_i - \bar{H}_i)^2/(2\sigma_i^2)]$, so that

$$N(t) = N_f \int g(H_{B_2A}) \exp[-(k_{B_2A}/N_f)t] dH + N_s \int g(H_{B_3B_2}) \exp[-(k_{B_3B_2} \times N_f)t] dH$$

The temperature dependence of the rate coefficients (above ~40 K) is governed by the Arrhenius relation, $k_i = A_i \exp[-H_i/(RT)]$, with pre-exponential A_i and enthalpy H_i of barrier i ; R is the gas constant and T the absolute temperature. The parameters A_i , $\bar{H}_i$, σ_i for k_{B_2A} and $k_{B_3B_2}$ were determined from a fit with the above equation and are compiled in the inset of Fig. 4c. Also included are the Arrhenius parameters for $k_{B_2B_3}$, determined from the temperature dependence of the population ratio N_f/N_s , using $k_{B_2A}(\bar{H}_{B_2A})$. If return from the B₃ avoids state B₂, the kinetics will be governed by a rate coefficient k_{B_3A} , with Arrhenius parameters essentially identical to those of $k_{B_3B_2}$.

Data collection and refinement

X-ray diffraction data sets were collected with a Siemens HISTAR multiwire proportional counter using CuKα radiation from an ENRAF NONIUS FR591 rotating anode generator. The intensities of the reflections were integrated and scaled with the program SAINT (Siemens). Subsequent data processing was performed with the CCP4 programs AGROVATA and TRUNCATE²⁹. The refinement was done using the program X-PLOR 3.1 (ref. 30). No angle restraints were used for the CO molecule and the proximal histidine H93 with respect to the haem iron. The bond restraints for the CO molecule and H93 as well as the pyrrole nitrogen bonds to the iron atom were weakened from the standard X-PLOR values to allow for an almost unrestrained refinement. During the course of the refinement, the models were verified by calculating simulated annealing omit maps for the regions of interest.

Photolysis of crystals

Crystals were illuminated from opposite sides with light from a continuous-wave, frequency-doubled Nd-YAG laser (532 nm). At the laser powers needed, substantial heating of the crystal cannot be avoided. We took advantage of this fact and used the photolysis laser power for supplemental heating of the crystal while the cryostream temperature control was set at 105 K. A temperature calibration was achieved by noticing that Debye rings started to develop at a photolysis laser power of 200 mW, reflecting a phase transition in the mother liquor that occurs in the dark at $T > 180$ K. To obtain the photoproduct below 180 K, we decreased the laser power from 100 to 30 mW during the time course of 20 h. Subsequently, data were collected under 30-mW illumination. To photolyse above 180 K, we started the illumination with a laser power of 200 mW instead. These illumination protocols were carefully developed using temperature-derivative

spectroscopy in the CO-stretch bands of L29W MbCO crystals to ensure that the maximum fraction of molecules would be trapped in the slow states²⁷.

Received 2 September 1999; accepted 4 January 2000.

- Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373 (1975).
- Steinbach, P. J. *et al.* Ligand binding to heme proteins: connection between dynamics and function. *Biochemistry* **30**, 3988–4001 (1991).
- Olson, J. S. & Phillips, G. N. Jr. Kinetic pathways and barriers for ligand binding to myoglobin. *J. Biol. Chem.* **271**, 17593–17596 (1996).
- Henry, E. R., Sommer, J. H., Hofrichter, J. & Eaton, W. A. Geminate recombination of carbon-monoxide to myoglobin. *J. Mol. Biol.* **166**, 443–451 (1983).
- Huang, X. & Boxer, S. G. Discovery of new ligand binding pathways in myoglobin by random mutagenesis. *Nature Struct. Biol.* **1**, 226–229 (1994).
- Quillin, M. L. *et al.* Structural and functional effects of apolar mutations of the distal valine in myoglobin. *J. Mol. Biol.* **245**, 416–436 (1995).
- Scott, E. E. & Gibson, Q. H. Ligand migration in sperm whale myoglobin. *Biochemistry* **36**, 11909–11917 (1997).
- Parak, F., Knapp, E. W. & Kucheida, D. Protein dynamics: Mössbauer-spectroscopy on deoxy-myoglobin crystals. *J. Mol. Biol.* **161**, 177–194 (1982).
- Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron-scattering. *Nature* **337**, 754–756 (1989).
- Nienhaus, G. U., Heinzl, J., Huenges, E. & Parak, F. Protein crystal dynamics studied by time-resolved analysis of X-ray diffuse scattering. *Nature* **338**, 665–666 (1989).
- Rasmussen, B. E., Stock, A. M., Ringe, D. & Petsko, G. A. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357**, 423–424 (1992).
- Schlichting, I., Berendzen, J., Phillips, G. N. Jr & Sweet, R. M. Crystal structure of photolysed carbonmonooxy-myoglobin. *Nature* **371**, 808–812 (1994).
- Teng, T. Y., Srajer, V. & Moffat, K. Photolysis-induced structural changes in single crystals of carbonmonooxy myoglobin at 40 K. *Nature Struct. Biol.* **1**, 701–705 (1994).
- Hartmann, H. *et al.* X-ray structure determination of a metastable state of carbonmonooxy myoglobin after photodissociation. *Proc. Natl Acad. Sci. USA* **93**, 7013–7016 (1996).
- Lim, M., Jackson, T. A. & Anfinsen, P. A. Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin. *Nature Struct. Biol.* **4**, 209–214 (1997).
- Srajer, V. *et al.* Photolysis of the carbon-monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science* **274**, 1726–1729 (1996).
- Vitkup, D., Petsko, G. A. & Karplus, M. A comparison between molecular dynamics and X-ray results for dissociated CO in myoglobin. *Nature Struct. Biol.* **4**, 202–208 (1997).
- Nienhaus, G. U., Mourant, J. R. & Frauenfelder, H. Spectroscopic evidence for conformational relaxation in myoglobin. *Proc. Natl Acad. Sci. USA* **89**, 2902–2906 (1992).
- Springer, B. A., Sligar, S. G., Olson, J. S. & Phillips, G. N. Jr. Mechanisms of ligand recognition in myoglobin. *Chem. Rev.* **94**, 699–714 (1994).
- Tilton, R. F., Kuntz, I. D. & Petsko, G. A. Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9-Å. *Biochemistry* **23**, 2849–2857 (1984).
- Hirota, S. *et al.* Perturbation of the Fe–O₂ bond by nearby residues in heme pocket: Observation of ν_{Fe-O_2} Raman bands for oxymyoglobin mutants. *J. Am. Chem. Soc.* **118**, 7845–7846 (1996).
- Schoenborn, B. P., Watson, H. C. & Kendrew, J. C. Binding of xenon to sperm whale myoglobin. *Nature* **207**, 28–30 (1965).
- Elber, R. & Karplus, M. Enhanced sampling in molecular-dynamics: use of the time-dependent Hartree approximation for a simulation of carbon-monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* **112**, 9161–9175 (1990).
- Frauenfelder, H., Parak, F. & Young, R. D. Conformational substates in proteins. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 451–479 (1988).
- Nienhaus, G. U. & Young, R. D. Protein dynamics. *Encyclopedia Appl. Phys.* **15**, 163–184 (1996).
- Alben, J. O. *et al.* Infrared spectroscopy of photodissociated carboxymyoglobin at low temperatures. *Proc. Natl Acad. Sci. USA* **79**, 3744–3748 (1982).
- Nienhaus, G. U., Mourant, J. R., Chu, K. & Frauenfelder, H. Ligand binding to heme proteins: the effect of light on ligand binding in myoglobin. *Biochemistry* **33**, 13413–13430 (1994).
- Springer, B. A. & Sligar, S. G. High-level expression of sperm whale myoglobin in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **84**, 8961–8965 (1987).
- Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brünger, A. T. X-PLOR Version 3.1. (Yale Univ. Press, New Haven, CT, 1992).
- Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).

Acknowledgements

We thank J. S. Olson for providing the L29W plasmid, K. Nienhaus for the sample preparation, and E. Hausteiner for expert assistance with the kinetic experiments. This work was supported by DFG Sonderforschungsbereich 533 and Graduiertenkolleg 328, Stiftung Volkswagenwerk and the University of Ulm.

Correspondence and requests for materials should be addressed to G.U.N. (e-mail: uli@uiuc.edu). Crystallographic coordinates and structure factor amplitudes for all four data sets have been deposited in the Protein Data Bank under accession codes 1DO1, 1DO3, 1DO4 and 1DO7.

Protein concentration

This week's quota of new products concentrates on proteins and peptides.

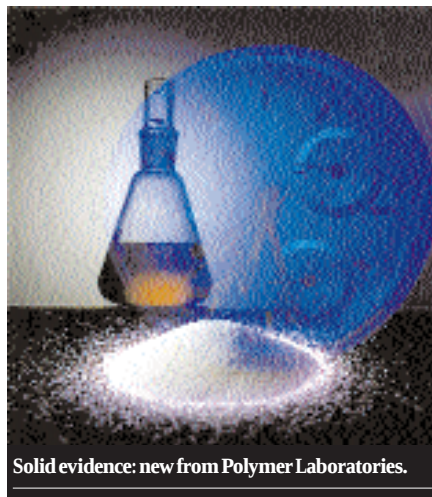
Solid-phase synthesis

Polymer Laboratories www.polymerlabs.com

More support for peptide synthesis

Three new resins from Polymer Laboratories, PL-FMP, PL-FDMP and PL-Rink Resin, have been introduced as aids in the synthesis of carboxamide compounds, important in peptide synthesis and combinatorial chemistry. The formyl-containing resins can also be used for the generation of sulphonamides. PL-FMP and PL-FDMP are both manufactured from copolymerized PL-CMS resin. They are available in a variety of particle sizes, and include higher loading versions for combinatorial chemistry. PL-Rink Resin (Fmoc Rink Amide AMS resin) is a tool for amide manufacture, particularly in peptide synthesis.

Reader Service No. 100



Solid evidence: new from Polymer Laboratories.

Endotoxin remover

Sterogene Bioseparations

www.sterogene.com

Plasma proteins come clean

ActiClean Etox is a solid-phase reagent with a high affinity for endotoxin, intended for the removal of endotoxin from human plasma proteins and derivatives, including serum albumin, intravenous IgG, and high-purity factor VIII and factor IX products. ActiClean Etox has an endotoxin binding capacity of 200–20,000 EU ml⁻¹, which varies with the application. Affinity of the immobilized ligand for proteins is very low and ActiClean Etox may be used either in a batch or on a chromatography column. The reagent is readily regenerated with 1 M NaOH, and sterilized by autoclaving. A Drug Master File is available for this product in the US and Europe.

Reader Service No. 101

Protein immobilizer

Exiqon

www.exiqon.com

Proteins and peptides halted in their tracks

Exiqon's anthraquinone photo-coupling immobilizers provide simple one-step procedures for binding most biomolecules to both natural and synthetic polymers. Immobilizer microplates and strips covalently bind low concentrations of proteins, peptides or DNA and can be used in colorimetric, bioluminescence, chemiluminescence and fluorescence assays. Typical applications include human *in vitro* and veterinary diagnostic use, functional and pharmacogenomics (HTS), analytical research, food safety and pesticide analysis. The coupling procedure does not require any preactivation steps involving toxic chemicals.

Reader Service No. 102

Catalogue

Hybaid

www.hybaid.com

The new catalogue, also available from sales@hybaid.com

Hybaid's latest catalogue is now available. It features over 200 new products including customized oligonucleotides, peptides, and gene synthesis from the Interactive Virtual Laboratory. Hybaid also claim to have one of the largest ranges of restriction endonucleases on the market. AutoPrep-12 is available for high-throughput extraction of plasmid DNA, DASH for SNP scoring, and the versatile biochip, XNA on Gold.

Reader Service No. 103

G-protein-coupled receptors

Packard Instrument Company

www.packardinst.com

A clear signal for screening

The SignalScreen range of over 60 cloned G-protein-coupled receptors is available from Packard through an arrangement with the company's BioSignal subsidiary. These cloned receptors have been purified and are ready for membrane preparation, with pharmaceutical screening programs as the main application.

Reader Service No. 104

Two-hybrid screening

Clontech Laboratories

www.clontech.com

Positive identification

The Matchmaker Two-Hybrid System 3 is a GAL4-based system for identifying and confirming protein interactions. The system features a new yeast strain that aims to eliminate false positives through the use of three reporters with distinct control regions. The new vectors include unique epitope tags,

different bacterial selection markers, and integrated T7 promoters to expedite independent confirmation of protein interactions by *in vitro* co-immunoprecipitation assays. This system comes with all the necessary reagents for two-hybrid screening, including sequencing primers, dropout supplements and Yeast-maker Carrier DNA.

Reader Service No. 105

Protein synthesis

Ambion

www.ambion.com

High-yielding and flexible

Two kits from Ambion, PROTEINscript II and PROTEINscriptPRO, are designed for the transcription of DNA into RNA, followed by the translation of the RNA produced into protein. These kits can be used to translate either eukaryotic or prokaryotic proteins and are said by Ambion to be among the highest-yielding and most flexible protein-synthesis kits available. Linear DNA, supercoiled DNA or unpurified PCR products can be used as the template for protein synthesis with these kits.

Reader Service No. 106

Criterion

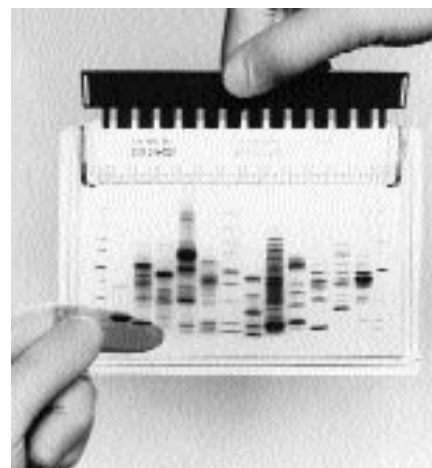
Bio-Rad

www.criterion.bio-rad.com

Mini in size, maxi in throughput

Bio-Rad's new precast gel system is an advanced mini gel electrophoresis outfit with a large sample capacity, making it possible to increase the number of results obtained with a single gel. This should be especially useful for high-throughput labs. The cell and cassette include features designed to eliminate buffer leaking and to make assembly simple and error-free.

Reader Service No. 107



Not a big band: Bio-Rad's mini system.



See the point? Millipore's ZipTip.

Sample preparation

Millipore www.millipore.com/ziptip
Zippy time saver

ZipTip Pipette Tips, for desalting femtomole-level volumes of peptides in under a minute, are now available containing C_4 resin for use with higher-molecular-weight peptides. ZipTip can be placed on a standard single or multi-channel pipettor, standard 22-gauge blunt-end HPLC needle or automated liquid handling/sample preparation station. A 0.6- μ l bed of either C_4 or C_{18} resin is fixed at the end of the ZipTip device and the sample is aspirated and dispensed through the resin to bind and wash peptides, proteins or oligonucleotides. ZipTips help save time in the preparation of samples for mass spectroscopy, capillary electrophoresis and other analytical techniques.

Reader Service No. 108

UV-Plates

Corning Science Products www.corning.com
The bottom line

These Costar brand 96-well UV-Plates boast UV-transparent bottoms to make them suitable for determining protein and/or nucleic acid concentrations. The UV-transparent

bottoms are moulded, rather than adhered, to an acrylic top for greater strength and maximum leak resistance. The plates are read down to 220 nm, and they give consistently low background and well-to-well uniformity. UV-Plates are designed for zero per cent crosstalk, are certified DNase and RNase free and are made from USP Class VI materials in an ISO 9002 registered facility.

Reader Service No. 109

Proteasome assay kit

Affiniti Research Products

www.affiniti-res.com & www.proteasome.com

Rapid determination of 20S proteasome activity

The basis of this assay for 20S proteasome activity *in vitro* is the measurement of the hydrolysis over time of the fluorogenic peptide substrate Suc-Leu-Leu-Val-Tyr-AMC by SDS-activated proteasome. This peptide substrate is cleaved by the chymotrypsin-like activity of the proteasome, releasing free AMC which can then be detected using a fluorimeter. The kit contains all buffers and reagents necessary for the analysis of 20S proteasome activity in cuvettes (sufficient reagents for 25 $\times$ 1 ml assays) or a 96-well microtitre plate (sufficient for 96-well-based assays). The kit may be used for qualitative and quantitative determination of 20S proteasome activity, 20S proteasome inhibition, 20S proteasome activation and 20S proteasome latent activity.

Reader Service No. 110

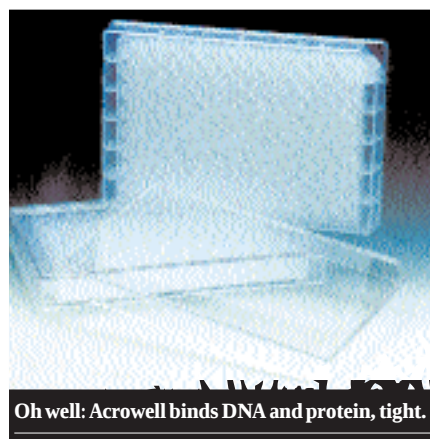
Acrowell

Pall Gelman Laboratory

www.pall.com/gelman

No crosstalk with high-binding plates

AcroWell filter plates contain BioTrace NT and BioTrace PVDF membranes to bind biomolecules such as protein and DNA. The upstream binding membrane layer is supported by hydrophobic PTFE. The lower mem-



Oh well: Acrowell binds DNA and protein, tight.

brane acts as a barrier to passive flow, allowing long incubations with the wells filled with hybridization and immunodetection solutions. AcroWell plates can be used to bind biomolecules for a variety of applications including ELISA, ELISPOT, kinase assays, nucleic acid hybridization, molecular diagnostics, cell-based immunodiagnostics and protein: protein and protein:nucleic acid interactions.

Reader Service No. 111

Protease inhibitors

Sigma-Aldrich

www.sigma-aldrich.com

The cocktail hour

Sigma-Aldrich's new ready-to-use, pre-mixed and pre-measured protease-inhibitor cocktails are convenient to use and offer time and cost savings in protein isolation. A range of cocktails is available for use with bacterial, mammalian, yeast, fungal and plant cell extracts, as well as for polyhistidine-tagged proteins. Importantly, all cocktails are assayed in their application setting as part of the extensive Sigma-Aldrich quality-control programme. Careful manufacturing and consistent composition guarantee performance in the laboratory — without the complex weighing, mixing, pH adjustment and special storage needed when preparing an inhibitor mix in-house.

Reader Service No. 112

Mini-catalogue

Alexis Biochemicals

www.alexis-corp.com

Read all about it

A new 12-page immunology and cancer research mini-catalogue is now available from Alexis Biochemicals. This mini-catalogue features new releases for apoptosis, CD markers, cell adhesion, cytokines and tumour biology research. New ELISA kits for p53, cathepsin L and leukaemia inhibitory factor, and a summary of kits for apoptosis detection are also included.

Reader Service No. 113

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

Check out our homepage at <http://www.resgen.com>

Peptides
\$14/residue
10 residue minimum

Polyclonal Antibodies - \$1,195

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 4

Lysyl Endopeptidase
(Achromobacter Protease 1)
Specifically cleaves all Lys-X bonds including Lys-Pro.
Highly Stable, Water Soluble

Wako

Wako Bio-Products
Wako Chemicals USA, Inc.
1000 Ball Road
Farmingdale, NY 11737
Telephone: (516) 271-7577
Fax: (516) 271-7701
(800) 922-6255
BioProducts@wako.com

READER ENQUIRY NO. 23